University of
Sheffield

Thaislate: an AI-Enhanced Tool for Thai-English Grammar Learning

Papatchol Thientong

Supervisor: Ms. Varvara Papazoglou

A report submitted in fulfilment of the requirements for the degree of MSc in Computer Science

in the

School of Computer Science

September 4, 2025

Declaration

All sentences or passages quoted in this report from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s). Any illustrations that are not the work of the author of this report have been used with the explicit permission of the originator and are specifically acknowledged. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure in this project and the degree examination as a whole.

Name: Papatchol Thientong

Signature: Papatchol Thientong

Date: 4 September 2025

Abstract

Thai learners struggle with English tenses because Thai lacks grammatical tense markers. Existing translation tools provide corrections without explanations, leaving learners unable to understand their errors. This dissertation develops Thaislate, a proof-of-concept system demonstrating how large language models can bridge this gap through context-aware grammar explanations.

The system integrates three specialised models in a pipeline architecture: Typhoon Translate 4B for Thai-English translation, custom-trained XLM-RoBERTa achieving 94.7% accuracy on 24-category tense classification, and Typhoon 2.1 12B generating educational explanations. User testing with 38 Thai learners produced 474 ratings averaging 4.2/5, with explanation quality rated 4.33/5 despite 74% pipeline tense classification accuracy. The disconnect between technical performance and user satisfaction validates the approach: learners value clear explanations even when imperfect. The system successfully serves real users with 99.5% uptime, establishing technical feasibility and user acceptance as foundation for future longitudinal studies on learning effectiveness.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Ms. Varvara Papazoglou, for her invaluable guidance, patience, and expertise throughout this research project. Her willingness to take interest in supervising my self-proposed project on Thai-English grammar learning demonstrates remarkable openness to interdisciplinary research. Her insightful feedback and unwavering support made this dissertation possible, and her remarkable ability to fill our scheduled one-hour weekly meetings with engaging discussion for nearly the full duration deserves particular recognition.

Special thanks are also due to my loyal companion, Oteung, my Labrador back home in Thailand, who has been patiently waiting for my return while I pursued academic adventures abroad. Though separated by continents and unable to appreciate the finer points of XLM-RoBERTa architecture, Oteung's steadfast loyalty and presumed excitement about my eventual homecoming have provided motivation during the more challenging moments of this research. I promise the extended walks we'll take together will more than compensate for the time spent debugging Flask applications.

Finally, I acknowledge the 218 Thai English learners who participated in the user study, whose insights were instrumental in validating the design decisions and educational effectiveness of the Thaislate system.

Contents

List of Figures

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Imagine a Thai university student named Somchai working on his thesis proposal in English. He writes "I am study at university for three years" and knows something sounds wrong, but cannot explain why. His translation app suggests "I have been studying at university for three years," but offers no explanation of why the present perfect continuous tense is appropriate here. Frustrated, Somchai memorises the correction without understanding the underlying temporal logic, leaving him likely to repeat similar errors.

This scenario illustrates a widespread challenge: Thai learners struggle with English tense usage because Thai is a "tenseless language" that relies on contextual cues rather than grammatical markers to convey time relationships. Current educational tools compound this problem. Translation systems provide corrections without explanation, while grammar instruction emphasises rote memorisation over contextual understanding.

To address these challenges, this project investigates a central research question: **Can AI-powered tools help Thai learners understand English tense usage through contextual, automatically-generated explanations?** This question drives the development of "Thaislate", a web-based tool that helps Thai learners understand English tenses through explanations in their native language. When a user types a Thai sentence like "chan gin khao chao laew" (I ate breakfast already), the system not only translates it to "I have eaten breakfast" but also explains *why* the present perfect tense is used instead of past simple. The explanation clarifies that the Thai word "laew" (already) indicates a completed action with present relevance, which corresponds to English present perfect usage. The tool acts like a knowledgeable tutor who understands both languages and can bridge the conceptual gap between Thai and English temporal systems, transforming simple translation interactions into learning opportunities.

1.2 Aims and Objectives

This research, motivated by the central research question, develops Thaislate as a proof-of-concept system that demonstrates how large language models can be integrated to provide grammar-aware translation with educational explanations. The research validates the technical

approach and user acceptance, establishing the foundation for future studies on educational effectiveness. This aim evolved through iterative development, where initial goals were refined based on practical discoveries about model capabilities and educational requirements.

To achieve this aim, the project pursues five specific objectives that emerged from the development process and combine technical innovation with educational validation:

1. **Build the Core System:** To implement a three-component pipeline that emerged from experimental work, integrating Thai-English translation, tense identification, and educational explanation generation. This architecture was discovered through iterative testing when monolithic approaches failed to achieve acceptable accuracy, particularly in tense classification.

2. **Develop Tense Recognition System:** To create a hierarchical classification system through extensive manual dataset curation, capable of identifying not only broad tense categories (Past, Present, Future) but also detailed temporal distinctions (24 specific uses such as "completed action with present relevance" or "scheduled future events"). This objective became critical when instruction models proved unreliable at tense classification, necessitating a dedicated solution.

3. **Create an Intuitive Learning Interface:** To develop a user-friendly web application where Thai learners can input sentences in Thai language and receive accurate English translations accompanied by clear explanations that bridge Thai and English temporal concepts.

4. **Validate System Design and User Acceptance:** To assess both the technical accuracy of the system (through automated performance metrics) and user acceptance (through user studies with Thai English learners), validating that the tool addresses identified needs and establishing a foundation for future effectiveness studies.

5. **Contribute to Educational AI Research:** To contribute new insights into how artificial intelligence systems can be designed for educational applications in contexts where linguistic resources are limited, providing a model for similar cross-linguistic learning challenges.

1.3 Overview of the Report

This dissertation follows a logical progression from understanding user needs through system development to comprehensive evaluation, structured as follows:

Chapter 2: Literature Review examines existing research in machine translation for Thai-English language pairs, grammatical tense classification systems, and computer-assisted language learning tools. The chapter identifies gaps in current approaches and establishes the theoretical foundation for the proposed solution.

Chapter 3: User Study and Requirements Analysis presents empirical research with 218 Thai English learners to understand their specific challenges and preferences for grammar learning tools. Based on these findings, the chapter derives system requirements

that ensure the technical solution addresses real educational needs rather than theoretical assumptions.

Chapter 4: Design and Implementation - Core Models and Functionality focuses on the design and implementation of the three specialised AI models that form the system's educational core. The chapter presents the development journey from initial experimentation to final architecture, covering translation model selection (Typhoon Translate 4B), custom tense classification system (XLM-RoBERTa), and pedagogical explanation generation (Typhoon 2.1 12B), along with confidence-aware response strategies.

Chapter 5: Design and Implementation - Pipeline and Website describes how the individual models integrate into a complete educational system. The chapter covers preprocessing pipeline implementation (fragment detection, language validation, content filtering), web application architecture with user interface design based on empirical findings, analytics and feedback systems, and production deployment strategies that enable user testing and validation.

Chapter 6: Evaluation and Results presents comprehensive evaluation across three dimensions: technical performance of individual models, pipeline integration accuracy, and user acceptance validation. The chapter includes systematic testing of translation quality (93.2% fluency), classification accuracy (94.7% isolated, 74% pipeline), explanation quality (84.9% correctness), and user feedback from 38 active participants providing 474 ratings with mean scores exceeding 4.0/5 across all evaluation criteria, demonstrating strong user acceptance of the system design.

Chapter 7: Conclusion and Future Work concludes the dissertation by summarising key achievements, reflecting on major challenges encountered, and proposing future research directions. The chapter presents the project's contributions to educational NLP (integrated pipeline architecture, hierarchical tense classification, Thai-specific pedagogical approach), acknowledges technical and methodological challenges faced during development, and outlines pathways for advancing educational NLP systems including longitudinal learning studies, expanded grammatical coverage, and potential adaptation to other language pairs.

1.4 Relationship to Degree Programme

This project directly applies the core competencies of the MSc Computer Science with Speech and Language Processing programme. The Text Processing and Natural Language Processing modules provided essential foundations for implementing transformer-based classification and handling Thai-English cross-lingual processing, while the programme's evaluation methodologies guided the assessment framework. Although the interdisciplinary nature required extending beyond the curriculum to incorporate pedagogical principles, this integration demonstrates the programme's goal of preparing graduates to apply computational linguistics to real-world educational challenges.

Chapter 2

Literature Review

This review examines three interrelated research domains critical to building educational translation tools for Thai–English language pairs: machine translation for low-resource languages, hierarchical grammatical classification systems, and computer-assisted language learning (CALL). The convergence of these areas supports the development of advanced systems that integrate accurate translation with pedagogically grounded grammatical explanations. Recent literature from 2020 onwards is synthesised to identify core technical approaches, their relevance to Thai–English contexts, and implications for educational applications.

2.1 Machine Translation for Low-Resource Languages

2.1.1 Multilingual neural machine translation architectures

Multilingual neural machine translation (MNMT) offers substantial benefits for low-resource language pairs by enabling parameter sharing and cross-lingual transfer. The BLEU (Bilingual Evaluation Understudy) score is an automatic metric for evaluating machine translation quality by comparing n-gram overlaps between machine output and reference translations, with scores ranging from 0 to 1 (often reported as 0-100 when multiplied by 100) where higher values indicate better translation quality[1].

Ngo et al. [1] demonstrated this by training many-to-one multilingual Transformer[2] models, showing BLEU score improvements of +1.62 and +2.54 points for French–Vietnamese and English–Vietnamese translation over bilingual baselines. Their approach leveraged monolingual data to address data sparsity challenges inherent in low-resource settings.

This approach is directly relevant to Thai–English translation, where parallel data is scarce. Thai's morphological complexity, including the lack of explicit word boundaries, tonal system, and absence of grammatical tense markers, creates particular challenges for translation systems trained primarily on languages with different linguistic structures. MNMT can leverage related translation tasks to improve performance through cross-lingual learning.

[1]BLEU is calculated as: BLEU = BP $\cdot$ exp($\sum_{i=1}^{n} w_i \log p_i$), where BP is the brevity penalty to penalise overly short translations, w_i are weights (typically 1/n), and p_i is the precision of i-grams measuring how much of the candidate matches reference phrases.

[2]Transformers are neural network architectures that use self-attention mechanisms to process sequences in parallel rather than sequentially, enabling more efficient training and better capture of long-range dependencies.

2.1.2 Thai-specific language model development

The Typhoon series, introduced by Pipatanakul et al. [2, 3], marks a major step forward in Thai large language models (LLMs). Typhoon-7B was adapted from the English-centric Mistral-7B[3] via continual pretraining on a balanced mix of Thai and English text, transferring general world knowledge to Thai-specific tasks. Its training addressed key linguistic features such as the absence of explicit word boundaries and unique Thai script, supported by a custom tokeniser (a component that breaks text into smaller units for processing) achieving 2.62× greater efficiency than GPT-4 for Thai tokenisation through the addition of 5,000 Thai-specific tokens.

Evaluations on ThaiExam [2], a benchmark drawn from national high-school and professional certification exams, showed Typhoon's Thai understanding to be comparable with GPT-3.5, outperforming it on 4 out of 8 evaluation datasets. Typhoon 2 [3] expanded into a multimodal family (1B–70B parameters) with improved data filtering, enhanced instruction-following, and specialised capabilities for Thai document and speech processing. These features provide a strong foundation for educational translation and grammar learning applications.

2.1.3 Cross-lingual knowledge transfer and domain adaptation

Cross-lingual transfer techniques have emerged as a powerful solution for improving translation quality in low-resource settings. Gupta et al. [4] demonstrated the effectiveness of Cross-Lingual Knowledge Distillation (CLKD) for Answer Sentence Selection (AS2) tasks in multilingual question answering systems. AS2 involves ranking candidate answer sentences according to their probability of correctly answering a given question - a core component of retrieval-based QA systems. Their CLKD approach trains student models in low-resource languages (such as Bengali, Finnish, and Swahili) to mimic strong English teacher models without requiring labeled target language data.

The method works by having both teacher and student models process parallel question-answer pairs in their respective languages (source and target), with the student learning to replicate the teacher's probability distributions through KL-divergence loss. They evaluated CLKD on two newly created multilingual AS2 datasets: Xtr-WikiQA (10 languages including Arabic, Hindi, and Japanese) and TyDi-AS2 (8 typologically diverse languages).

Notably, CLKD achieved performance that rivals or exceeds supervised fine-tuning, with improvements being particularly pronounced for smaller models (e.g., mBERT-Base) and monolingual training scenarios. For instance, on the Xtr-WikiQA dataset, CLKD with XLM-RoBERTa-Large achieved P@1 scores[4] exceeding 80% for most target languages, outperforming supervised baselines by 5-10 percentage points. This finding is especially relevant for educational applications where computational efficiency and real-time response are critical, as it enables deployment of accurate multilingual QA systems without extensive

[3]Mistral-7B is an open-source large language model with 7 billion parameters developed by Mistral AI, known for its efficiency and strong performance despite its relatively small size compared to models like GPT-4.

[4]P@1 (Precision at 1) measures whether the top-ranked answer is correct, a key metric for ranking tasks where the system must identify the best answer from multiple candidates.

labeled data collection for each target language.

Complementing this work, Goyle et al. [5] conducted a comprehensive analysis of neural machine translation strategies for low-resource languages, evaluating techniques including back-translation, transfer learning, and focal loss on the mBART-CC25 model (a multilingual sequence-to-sequence model pre-trained on 25 languages). Focal loss is a modified loss function that down-weights easy examples and focuses training on hard negatives, particularly useful for addressing class imbalance. Their experiments on Sinhala, Nepali, Khmer, and Pashto demonstrated that back-translation with focal loss achieved state-of-the-art results, improving BLEU scores by 0.57 and 1.29 points for Sinhala-English and Nepali-English respectively. The study revealed that focal loss addresses the challenge of low-frequency token generation, a crucial consideration for grammatical constructions that may appear infrequently in training data. Furthermore, their transfer learning experiments showed that parent models trained on related high-resource languages can effectively initialise child models for low-resource pairs, though data quality proved more influential than methodology in some cases.

Together, these studies establish that cross-lingual knowledge transfer combined with domain-specific adaptation can significantly enhance translation systems for low-resource languages, offering both improved accuracy and computational efficiency, both key requirements for educational technology applications.

2.2 Cross-lingual Text Classification and Hierarchical Approaches

2.2.1 XLM-RoBERTa for multilingual classification

Multilingual transformer models have transformed cross-lingual text classification. Conneau et al. [6] introduced XLM-RoBERTa, a RoBERTa variant pretrained on 2.5 TB of CommonCrawl data across 100 languages. Its zero-shot transfer capability has delivered state-of-the-art results across diverse language pairs, with particularly strong performance on low-resource languages, improving 15.7% in XNLI accuracy[5] for Swahili and 11.4% for Urdu over previous XLM models. This makes XLM-RoBERTa an ideal foundation for Thai-English tense classification tasks.

While Karamanolakis et al. [7] demonstrated that minimal resources (just 20 translated seed words) can enable effective cross-lingual classification through their CLTS (Cross-Lingual Text Selection) method, achieving up to 12% accuracy improvements over multilingual BERT approaches, their focus on document-level classification differs from the sentence-level grammatical analysis required for tense classification. Nevertheless, their finding that task-specific seed words can guide classification is relevant: grammatical markers and temporal expressions could serve as similar anchor points for tense classification in Thai-English translation contexts.

[5]XNLI (Cross-lingual Natural Language Inference) is a benchmark dataset for evaluating cross-lingual understanding. Models are tested on their ability to determine whether sentence pairs exhibit entailment, contradiction, or neutral relationships across 15 languages. The accuracy metric represents the percentage of correctly classified premise-hypothesis pairs, with state-of-the-art models typically achieving 70-85% accuracy.

2.2.2 Hierarchical text classification methodologies

Hierarchical classification is well-suited to tasks with naturally structured label taxonomies. Mekala et al. [8] introduced the Coarse2Fine framework at EMNLP 2021, demonstrating how to perform fine-grained classification on coarsely-annotated data without fine-grained human annotations. Their approach uses label-conditioned fine-tuning with generative language models and incorporates hierarchy-aware regularisation objectives based on coarse-fine label constraints.

The framework was evaluated on two real-world hierarchical datasets. The New York Times (NYT) dataset consists of news articles classified into 5 coarse-grained genres (e.g., arts, sports) and 25 fine-grained categories (e.g., movies, baseball). The 20 Newsgroups (20News) dataset contains newsgroup documents partitioned into 6 broad groups (e.g., recreation, computers) and 20 fine-grained classes (e.g., graphics, windows, baseball, hockey). Notably, the researchers excluded three miscellaneous labels from 20News (misc.forsale, talk.politics.misc, talk.religion.misc) as their label surface names lacked focused meaning, which would interfere with the label-based weak supervision approach.

The framework achieved 92.62% Micro-F1 on the NYT dataset, significantly outperforming zero-shot baselines by leveraging the hierarchical relationship between coarse and fine labels through iterative weak supervision and bootstrapping.

This coarse-to-fine approach provides a framework for grammatical tense classification, where broad temporal categories (Past, Present, Future) can be progressively refined into specific tense-aspect combinations. The methodology of using label surface names as weak supervision signals could be adapted for tense classification in translated text.

However, while hierarchical methods have proven effective for monolingual classification tasks, their application to analysing grammatical structures in machine-translated text remains unexplored. This gap is particularly significant for educational applications where understanding the rationale behind tense choices is as important as the classification itself.

2.2.3 Grammatical error correction with explanations

Recent research has emphasised augmenting grammatical corrections with pedagogical explanations to enhance learning outcomes. Fei et al. [9] introduced EXPECT, a large-scale dataset containing 21,017 instances annotated with evidence words and grammatical error types across 15 categories. Their experiments demonstrated that explainable GEC (Grammatical Error Correction) systems can serve as effective post-processing modules for existing GEC systems. Human evaluations showed that when applied to GECToR (a sequence tagging approach to grammatical error correction) outputs, 82.4% of corrections received explanations, with 84.5% of these explanations deemed helpful for understanding the corrections.

Building on this interpretability focus, Qorib et al. [10] developed MoECE (Mixture of Error Correction Experts), which employs specialised expert networks for different error types. Their approach achieves state-of-the-art performance, improving F0.5 scores[6] by up to 1.32 points on BEA-2019[7] while maintaining interpretability through error type identification during inference. The model's routing mechanism learns to allocate tokens to appropriate experts based on error types, with analysis revealing that experts develop specialisations (e.g., achieving 83.7% accuracy for punctuation corrections).

These advances in explainable grammatical analysis demonstrate the value of providing learners with clear rationales for linguistic choices. However, current GEC systems focus on correcting errors in existing English text rather than explaining grammatical choices made during translation from a tenseless language. This leaves a significant gap in supporting learners who need to understand why certain tense choices are appropriate when translating from their native language, particularly when that language encodes temporal information through fundamentally different mechanisms.

2.3 Computer-Assisted Language Learning and Large Language Models

2.3.1 Large language models in educational applications

Large language models are increasingly applied in education for feedback generation, personalised content, and intelligent tutoring. Dong et al. [11] conducted a systematic review of 94 studies, finding that LLMs show potential in educational applications including content generation, serving as learning assistants, and supporting academic research. Their review provides crucial insights: the importance of maintaining pedagogical transparency to avoid over-reliance on AI, the need for culturally aware content generation particularly for non-English speakers, and the value of explanation-based approaches over simple correction. However, their review also highlights critical challenges, including risks to students' critical thinking development, academic integrity concerns, and ethical issues that require careful consideration before widespread implementation in educational settings.

These findings underscore that effective educational translation tools must go beyond accurate translation to provide transparent, culturally appropriate explanations that support learning rather than replacing understanding. For Thai learners, this means systems must explain not just what the correct English tense is, but why it appropriately captures the temporal and aspectual meaning expressed in the original Thai.

2.3.2 Generative AI adoption and learning outcomes in Thai contexts

While the potential of LLMs in education is well-documented, empirical evidence on their actual impact on learning outcomes remains limited. Waluyo and Kusumastuti [12] conducted

[6] $F_{0.5}$ is a variant of the F-measure that weights precision more heavily than recall, emphasising the importance of avoiding false positives in error correction.

[7] BEA-2019 refers to the Building Educational Applications 2019 shared task dataset, comprising 3,700 annotated texts from both native and non-native English writers across different proficiency levels, used as a standard benchmark for grammatical error correction systems.

a mixed-methods study with Thai university English learners (n=25) examining GAI tool adoption and its correlation with academic performance. Their findings reveal a critical paradox: despite high student acceptance across performance expectancy (M=3.66), effort expectancy (M=3.61), and actual usage behaviour (M=3.52), no significant correlation existed between GAI usage and GPA (ρ=-0.06, p=0.76).

This disconnect between enthusiasm and outcomes highlights fundamental challenges for educational GAI implementation. Students reported improved efficiency and confidence but also expressed concerns about over-dependency, with one participant noting that "using AI has many disadvantages" despite acknowledging its benefits. Teacher observations reinforced these concerns, citing difficulties in assessing genuine student progress and risks of superficial engagement with learning materials.

Particularly relevant for Thai-English translation systems are the specific challenges students identified with language-specific accuracy. Student testimonials revealed fundamental issues with current GAI translation capabilities: "If we use AI to write compositions and submit them to teachers without double-checking, it could potentially lead to confusion in the message as Thai is a complex language. Sometimes, AI may not accurately translate Thai language fully" (S15, in Waluyo & Kusumastuti, 2024). This observation directly validates the need for specialised Thai-English translation models rather than general-purpose GAI tools.

Furthermore, students recognised that effective AI usage requires foundational knowledge: "AI reduced work time and increased convenience, although it may make mistakes, so we must use it correctly with basic knowledge" (S14). This insight supports the pedagogical approach of providing grammatical explanations alongside translations, ensuring learners develop the "basic knowledge" necessary to critically evaluate AI-generated content. Another student's mixed assessment, "using AI has many disadvantages" despite acknowledging improved writing completeness (S9), underscores the tension between efficiency gains and genuine learning, a gap that explanation-augmented systems could address.

These findings directly inform the design requirements for educational translation tools. Rather than simply providing translations or corrections, systems must foster deep understanding to bridge the gap between tool usage and learning outcomes. For Thai-English contexts specifically, this suggests that grammatical explanations and pedagogical transparency, as emphasised by Dong et al. [11], are essential rather than optional features.

2.3.3 Multilingual capabilities and cross-lingual applications

Chan et al. [13] explored grammatical error correction for code-switched sentences[8], developing a novel method for generating synthetic CSW (Code-Switched) GEC datasets by translating spans within existing GEC corpora. Using XLM-RoBERTa as their base model with a sequence-tagging approach (GECToR), they investigated six different span selection methods, finding that the linguistically-motivated "noun-token" method yielded the best results. Their optimised model achieved an average F0.5 increase of 1.57 across English-Chinese (EN-CH), English-Korean (EN-KO), and English-Japanese (EN-JA) test sets while maintaining monolingual performance on the BEA-2019 dataset.

[8]Code-switching refers to utterances that mix two or more languages within a single discourse.

Notably, they discovered that models trained on one CSW language pair generalised relatively well to typologically similar languages; for instance, the English-Korean model slightly outperformed the English-Japanese model on English-Japanese data. This cross-lingual transferability is particularly relevant for Thai-English contexts where code-switching is common among bilingual speakers, suggesting that GEC systems can be adapted to handle mixed-language utterances without sacrificing accuracy on monolingual text, especially when linguistic similarities exist between the target languages.

2.3.4 Multi-Agent Systems for Language Learning Applications

Recent advances in multi-agent systems enable multiple AI agents with specialised roles to collaborate on complex educational tasks. Unlike traditional single-model approaches, multi-agent architectures divide tasks among specialised components that communicate through structured protocols, allowing each agent to focus on specific capabilities such as linguistic analysis, pedagogical strategy, or error detection.

Wu et al. [14] introduced AutoGen, a framework for orchestrating conversable agents through structured conversations. AutoGen addresses the challenge of coordinating multiple LLMs by implementing a conversation programming paradigm where developers define agents with specific capabilities and program their interactions using standard Python control flow rather than complex prompt engineering. This approach proved particularly effective for educational tasks: their math problem-solving application achieved 69.48% accuracy on the MATH dataset compared to 55.18% for GPT-4 alone, while reducing human intervention requirements by 3-10x. The framework's modular design allows different agents to handle distinct aspects of educational systems, such as content generation, validation, and pedagogical assessment, mirroring natural teaching workflows where specialists collaborate on curriculum development.

Complementing this approach, Khattab et al. [15] developed DSPy to address the brittleness of hand-crafted prompts in LLM pipelines. DSPy treats prompts as programmatic systems that can be automatically optimised through a compilation process. Rather than manually tuning prompts, developers specify high-level signatures (such as "question to answer" or "sentence and grammar rule to explanation"), and DSPy generates optimised prompts through bootstrapping with training examples. Their experiments demonstrated 25-65% performance improvements across various tasks, with automatically discovered prompt strategies outperforming expert-crafted alternatives.

These multi-agent and pipeline optimisation approaches directly address challenges in cross-linguistic education. Thai learners struggle with English tenses because Thai expresses time through context and aspect markers (laew for completion, gamlang for ongoing, ja for future) rather than verb conjugation. A multi-agent architecture can separate translation, tense classification, and explanation generation into specialised components, ensuring each stage is optimised for its specific task while maintaining pedagogical coherence. This modular approach enables systems to adapt explanations based on learner patterns while maintaining the flexibility to incorporate human feedback when needed.

2.4 Research Gaps and Future Directions

The literature review reveals several critical gaps that this dissertation addresses. Despite notable advances in individual areas, the integration of these technologies for Thai-English educational applications remains largely unexplored:

Limited Thai-specific educational NLP research – Most progress in NMT and CALL has not been tailored to Thai-English, whose tenseless structure and aspectual marking demand specialised solutions.

Hierarchical tense analysis of machine-translated text – While hierarchical classification methods exist for various NLP tasks, their application to analysing and explaining grammatical structures in machine-translated output remains unexplored. This is particularly crucial for educational contexts where learners need to understand not just what tense was chosen in translation, but why it appropriately captures the temporal meaning from the source language.

Integration gap between translation and grammatical explanation – Current systems either translate text OR provide grammatical feedback, but rarely integrate both capabilities in a pedagogically meaningful way. For Thai learners, understanding how their language's aspect markers and temporal context map to English tense choices requires systems that can both translate accurately and explain the grammatical rationale.

User-identified limitations in Thai-English GAI translation – Empirical evidence from Thai learners reveals that current GAI tools struggle with Thai's linguistic complexity, often producing translations that "lead to confusion in the message" [12], highlighting the need for specialised Thai-aware translation models with pedagogical explanations to support critical evaluation of AI output.

Adoption-outcome gap in educational GAI – Despite high acceptance and usage rates of GAI tools among Thai English learners, no correlation with improved academic outcomes has been demonstrated [12], suggesting current tools fail to foster the deep grammatical understanding necessary for genuine learning improvement.

Evaluation methodologies – Existing frameworks typically focus on technical metrics or isolated educational outcomes. More holistic evaluation combining translation accuracy, classification performance, explanation quality, and actual learning impact is needed for integrated educational NLP systems.

Chapter 3

User Study and Requirements Analysis

This chapter presents an analysis of user needs through an empirical study of Thai English learners, followed by the technical and pedagogical requirements derived from these findings. By grounding our requirements in actual user data, we ensure that the system design addresses real educational needs rather than theoretical assumptions.

3.1 User Study Design and Methodology

3.1.1 Research Objectives

The user study aimed to validate key assumptions about Thai English learners' challenges and preferences for grammar learning tools. The primary research questions were:

1. What specific challenges do Thai English learners face?
2. How do learners currently approach grammar learning, and what are their satisfaction levels with existing tools?
3. What features and explanatory approaches would be most valuable in a context-aware translation tool?
4. What is the target demographic and their technological readiness for AI-powered learning tools?

3.1.2 Survey Design and Implementation

A online survey was designed and distributed to Thai English learners through university networks and social media platforms. The survey consisted of five main sections:

1. **Demographics and Background**: Age, education level, English proficiency, and learning context
2. **Grammar Learning Challenges**: Specific difficulties with English grammar, particularly tense usage
3. **Current Tool Usage**: Experience with existing translation and grammar tools

4. **Feature Preferences**: Desired capabilities for an ideal grammar learning tool
5. **Technology Acceptance**: Comfort level with AI-powered educational tools

3.1.3 Ethical Considerations

The study received ethical approval from the University Research Ethics Committee of the University of Sheffield (Application Number 069760). All participants provided informed consent, and data was collected anonymously. The ethics application, ethics approval letter, participant information sheet, and consent form can be found in Appendix A.1, while the complete questionnaire used in the study is provided in Appendix B.

3.2 User Study Results

3.2.1 Participant Demographics

The survey received 234 initial responses, with 218 participants (93.2%) providing complete data after consent filtering as required by ethics approval. Participants were provided with both project information and consent forms before proceeding.

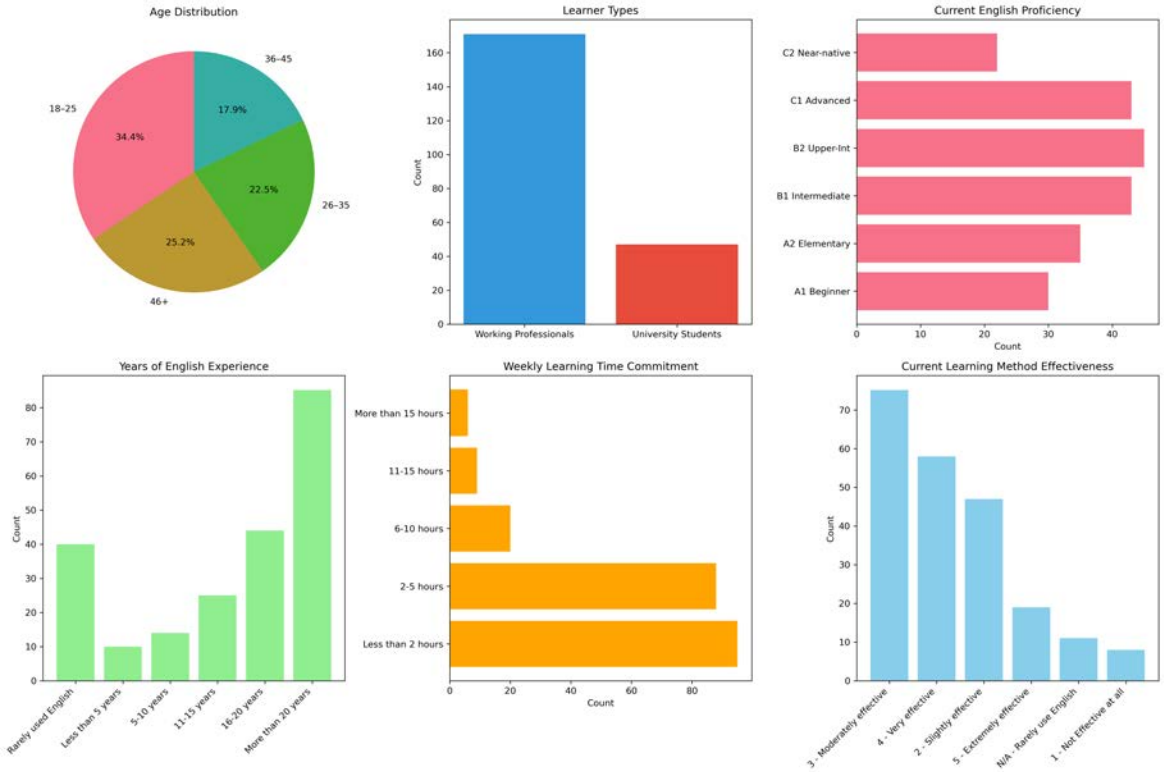


Figure 3.1: Participant Demographics and Learning Context (n=218)

Figure 3.1 reveals several key insights that shape our understanding of the target user base. The data shows that, the data challenges our initial assumption of a primarily university-focused tool. With working professionals comprising nearly 80% of respondents, the system

clearly addresses workplace learning needs as much as academic ones. This professional dominance suggests that English grammar improvement is a career-development priority for Thai adults.

The data reveals an interesting pattern: over 85 participants report more than 20 years of English experience, yet still identify various challenges when using the language. This distribution suggests that even extensive exposure to English does not necessarily eliminate all difficulties for Thai speakers, particularly in areas like grammar and tense usage that differ fundamentally between the two languages.

The effectiveness ratings reveal a critical gap in current learning methods. With most learners rating their approaches as only "moderately effective," there's clear demand for more effective alternatives. The concentration of learners at B1-B2 levels, combined with moderate satisfaction with existing methods, identifies an optimal intervention point where learners have sufficient foundation but still struggle with complex structures like tense usage.

3.2.2 Learning Challenges Analysis

This section analyses responses to three key survey questions:

- Question 12: “Which aspects of English do you find most challenging?” (multiple choice)
- Question 13: “How confident do you feel when using English in different situations?” (1-5 scale)
- Question 14: “Describe a specific situation where you had difficulty with English” (open text)

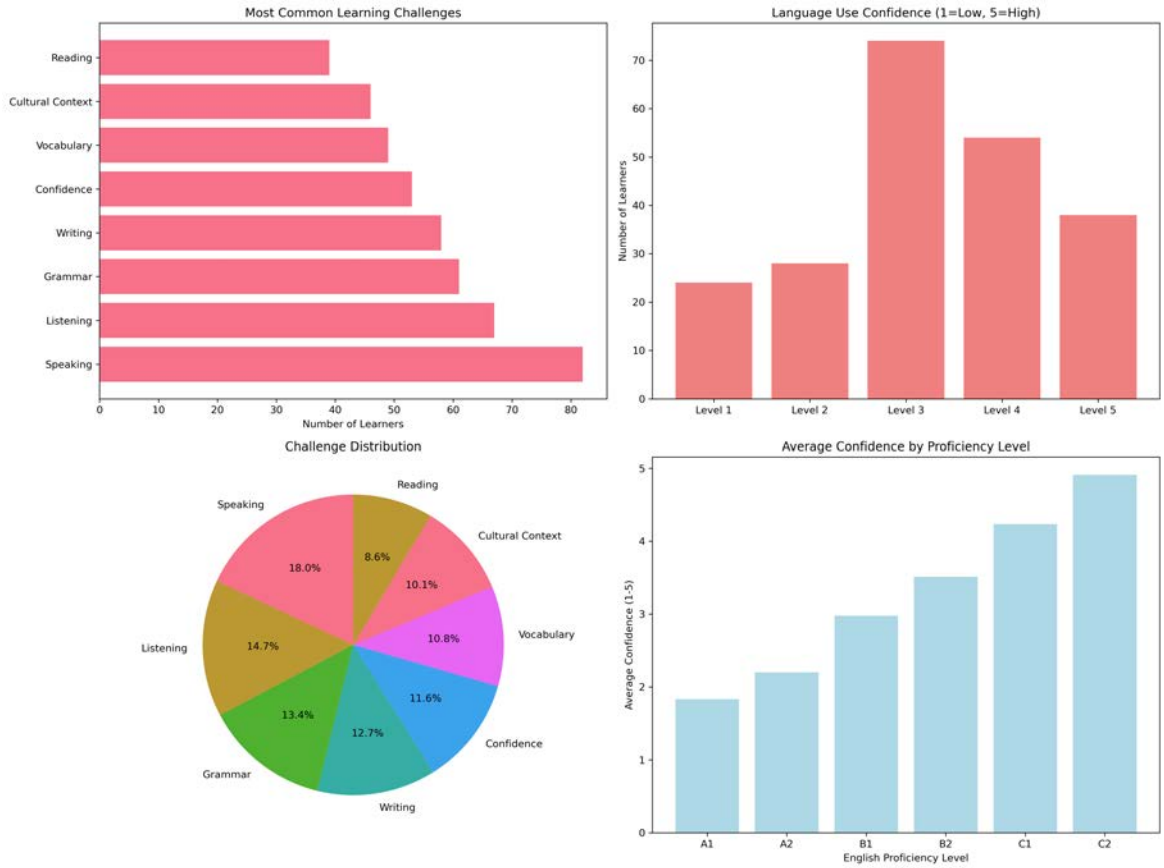


Figure 3.2: Learning Challenges and Confidence Analysis (n=218)

Figure 3.2 reveals critical insights for educational technology design. Speaking and Listening dominate as primary challenges (18.0% and 14.7% respectively), reflecting the inherent difficulties Thai learners face with English phonetics and real-time comprehension. These skills require interactive practice and immediate feedback that current automated tools struggle to provide effectively. Writing follows at 14.0%, representing another production-based skill requiring error detection and contextual understanding.

Grammar's position as the third-ranked challenge (13.4%) is important for system design. Unlike speaking or listening, grammatical analysis represents a challenge that computational approaches can address systematically through pattern recognition, rule application, and structured explanation generation. The relatively even distribution across all skill areas (vocabulary 11.7%, reading 10.5%, pronunciation 9.2%) indicates that learners face multifaceted challenges requiring integrated solutions rather than single-skill interventions.

The confidence distribution reveals patterns about learner self-perception. The concentration of learners at moderate confidence levels (Level 3 showing the highest frequency) suggests that most participants have moved beyond beginner uncertainty but have not reached higher proficiency. This "intermediate plateau" represents the optimal target audience for explanation-based tools: learners with enough foundation to understand grammatical concepts but still requiring guidance to advance further.

3.2.3 Learning Preferences and Support Needs

This section examines responses to questions about learning preferences:

- Question 24: "When you make a grammar mistake, what kind of feedback do you prefer?"
- Question 25: "How do you prefer to learn grammar rules?"
- Question 26: "How important is it to understand WHY a grammar rule works, not just HOW to use it?" (1-5 scale)

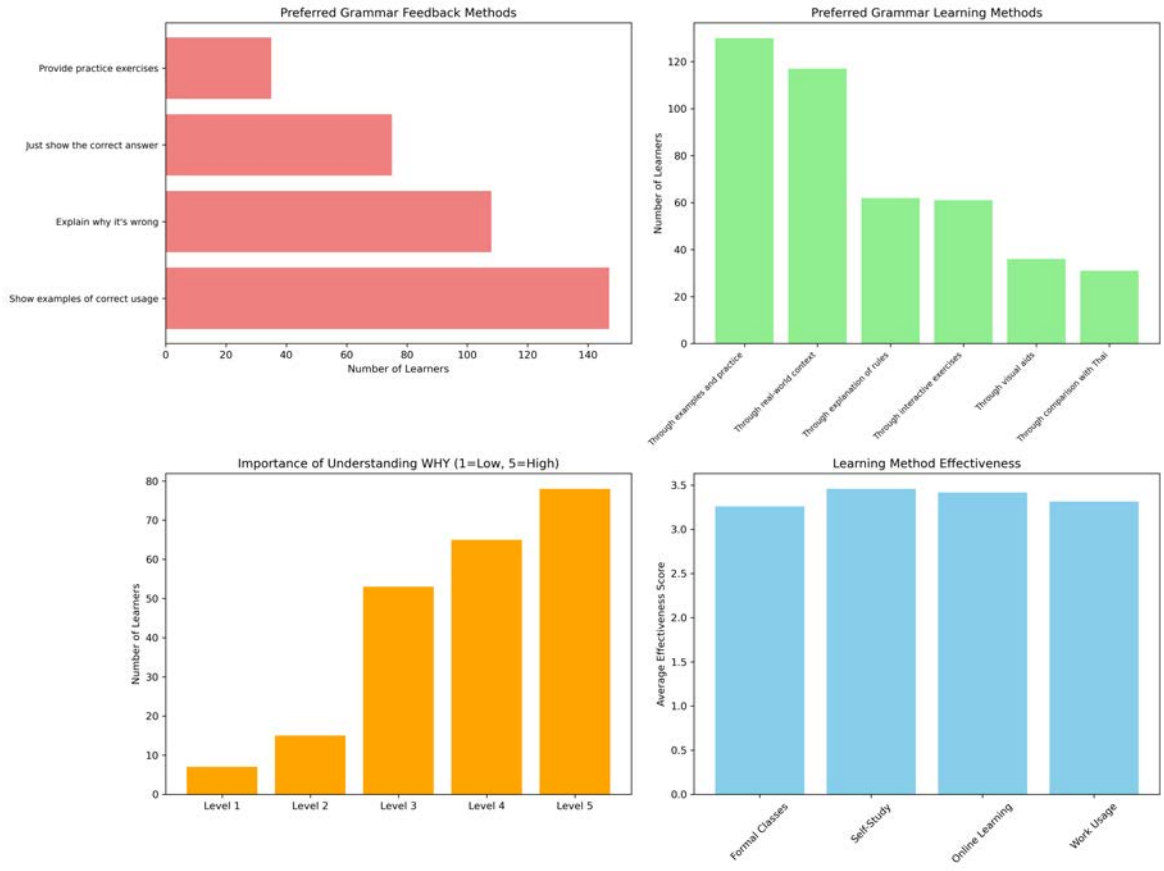


Figure 3.3: Grammar Learning Preferences and Support Methods (n=218)

Figure 3.3 reveals a fundamental shift in how Thai learners want to approach grammar learning. The most striking finding is the 2:1 preference ratio for "show examples of correct usage" versus "just show the correct answer." This suggests that learners want to understand patterns, not just memorise corrections.

The WHY importance ratings provide overwhelming validation for explanation-focused tools: with 160 out of 218 learners (73.4%) rating understanding "why" as high importance (Levels 4-5), and only 21 learners (9.6%) considering it low importance, the data strongly contradicts traditional grammar teaching approaches that focus on rote memorisation.

Perhaps most revealing is the uniform mediocrity of current learning methods. All approaches cluster around 3.2-3.5 effectiveness ratings, potentially indicating dissatisfaction rather than preference diversity. This "effectiveness plateau" suggests that existing methods, whether formal classes or self-study, are failing to meet learner needs adequately. The lack of a standout preferred method creates a clear opportunity for new approaches that combine contextual examples with explanatory depth.

The preference for "real-world content" and "examples-based learning" over "rule introduction" signals a desire for practical, applied grammar understanding rather than theoretical knowledge. This directly supports contextualised translation-based grammar instruction.

3.2.4 Technology Adoption and Feature Requirements

This analysis draws from survey questions about technology usage:

- Question 16: “How comfortable are you with using technology for learning?” (1-5 scale)
- Question 17: “How often do you use web-based learning tools?”
- Question 22: “Have you used any AI-powered tools before?” (Yes/No)

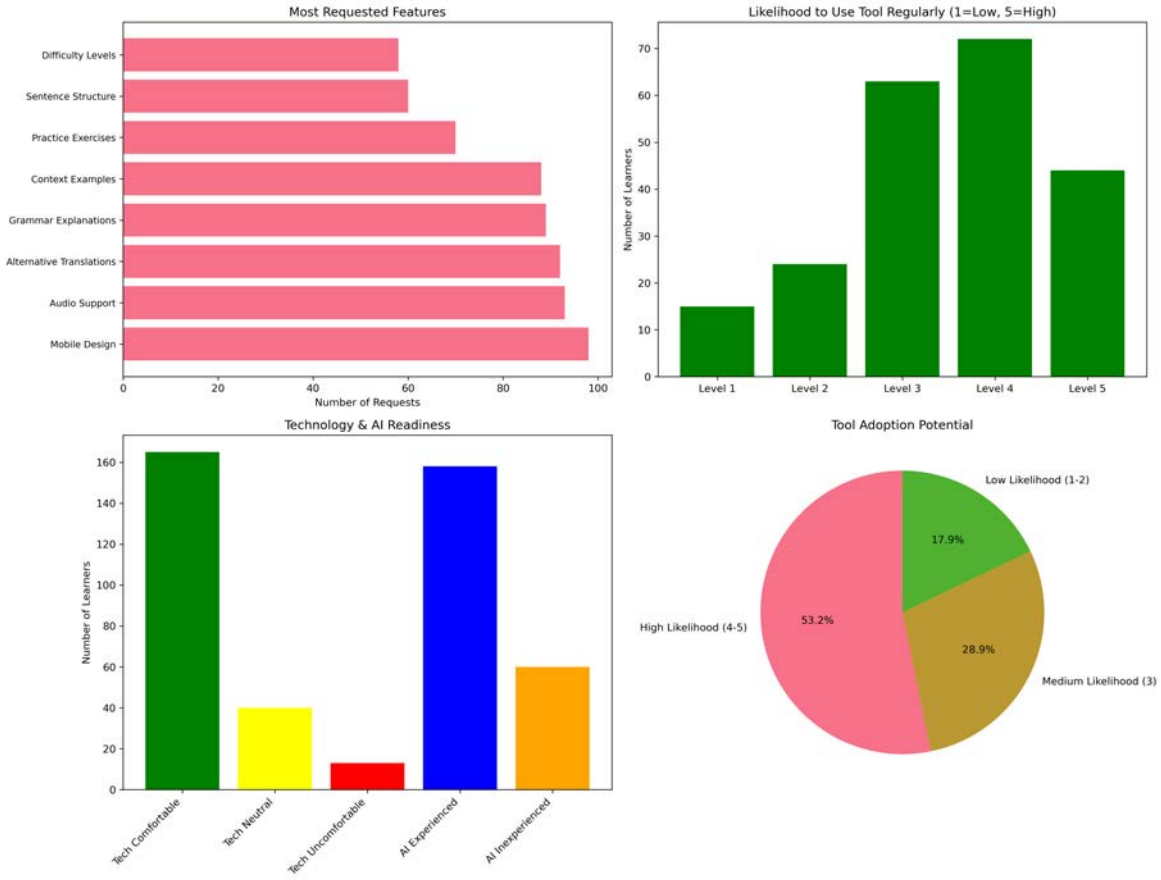


Figure 3.4: Technology Adoption Readiness and Feature Requests (n=218)

Figure 3.4 uncovers a crucial insight about modern language learning expectations: mobile design leads feature requests not because learners prefer phones over computers, but because it signals anywhere-anytime learning integration into busy professional lives. The clustering of requests around accessibility features (mobile design, audio support) suggests learners want tools that fit their lifestyle rather than requiring dedicated study time.

The technology readiness findings reveal a generational shift in Thai English learning. With 165 out of 218 participants (75.7%) having AI experience, Thai learners are more technologically experienced than traditional grammar tools assume. This high AI familiarity suggests that learners would not only accept but expect adaptive responses rather than static rule presentations.

Most encouraging for system viability is the adoption likelihood pattern: a strong majority of participants show high likelihood to use such tools regularly. This isn't just positive feedback; it represents genuine behavioural intent from learners who are already actively seeking better grammar solutions.

The feature prioritisation reveals strategic insights: "Alternative Translations" ranking highly suggests learners want to understand multiple ways to express ideas, not just correction of errors. This points toward an exploratory rather than corrective learning mindset, supporting translation-based grammar instruction that shows possibilities rather than just fixes problems.

3.2.5 Correlation Analysis and Key Relationships

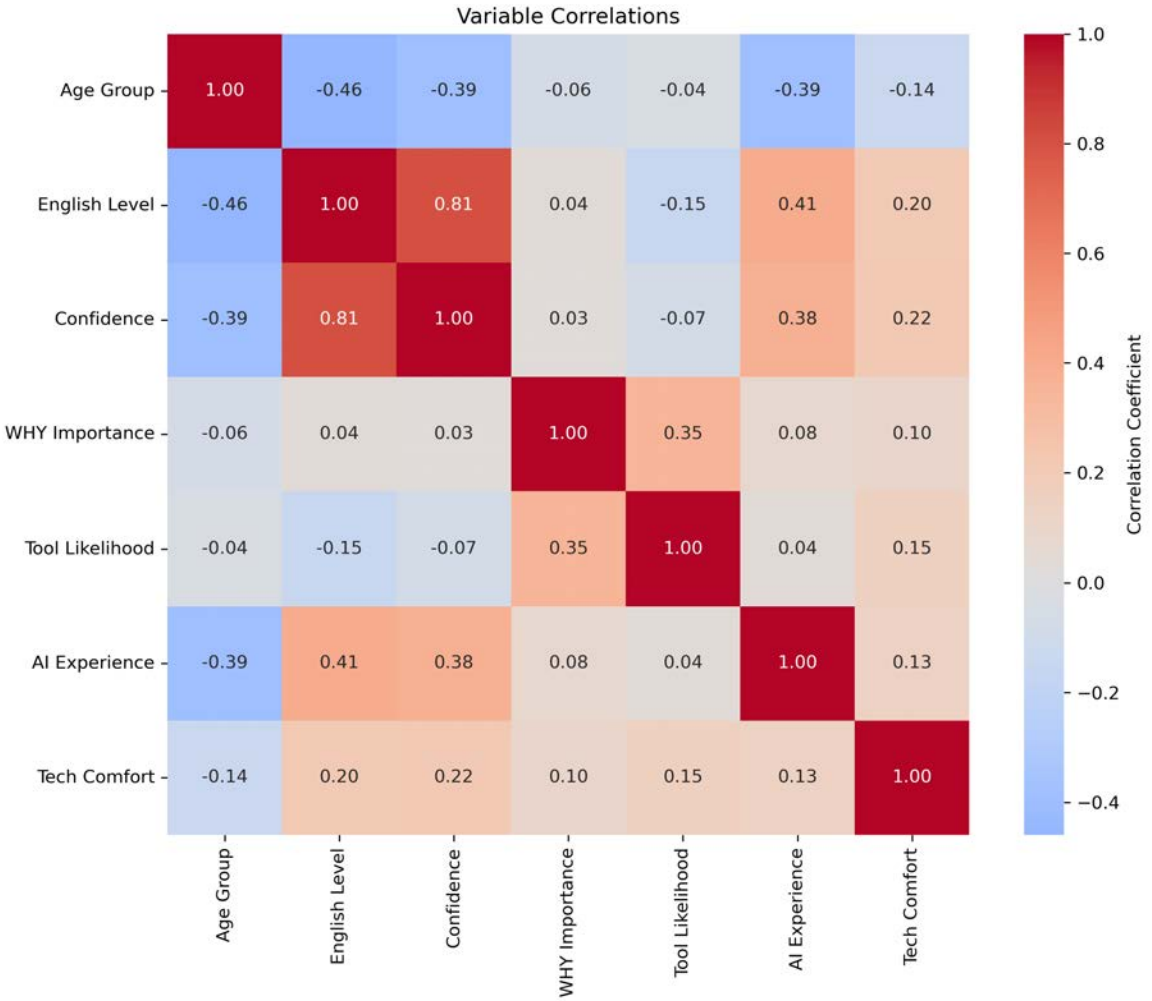


Figure 3.5: Variable Correlations and Statistical Relationships (n=218)

Figure 3.5 reveals the fundamental driver of learning tool success: the exceptionally strong English Level-Confidence correlation (0.81) indicates that confidence is not just a byproduct of learning but is intrinsically linked to perceived proficiency. This suggests that tools building genuine understanding (rather than providing quick fixes) may be more effective at improving learners' overall language confidence.

The most strategically important finding is WHY Importance's weak correlations across all demographics (all <0.35). This pattern indicates that the desire for explanatory learning is a fundamental learning preference rather than a demographic characteristic. Unlike features that might appeal to specific age groups or proficiency levels, explanation-focused approaches have universal appeal, making them ideal for broad-market educational tools.

Another revealing insight emerges from the age-related correlations: younger learners demonstrate higher English proficiency levels (age-proficiency correlation: -0.46) and greater confidence (age-confidence correlation: -0.39), likely reflecting improved educational standards

and earlier exposure to English in modern Thai education. These younger learners also show higher AI experience (age-AI correlation: -0.39), suggesting a generation that is both linguistically capable and technologically adept. This convergence presents an ideal user base: learners with strong foundational skills who are comfortable with AI-powered tools and seeking to refine their already substantial English abilities through deeper grammatical understanding.

The Tool Likelihood correlation with WHY Importance (0.35) reveals a crucial user segmentation insight: learners who value explanations are exactly those most likely to adopt new learning tools. This correlation suggests that explanation-focused positioning would naturally attract the most engaged and adoption-ready user segment, creating a self-reinforcing user base of committed learners.

3.2.6 Implications for System Design

The user study findings converge on clear design priorities that directly inform our system architecture and feature decisions. The data reveals not a population needing remedial support, but rather learners seeking deeper understanding of grammatical nuances. This critical distinction shapes our entire approach.

Strategic Focus on Grammar

While Speaking and Listening dominate learner challenges (18.0% and 14.7%), these require real-time interaction and pronunciation feedback that current automated systems may not adequately provide. Grammar, ranking third at 13.4%, represents the sweet spot where computational approaches excel. Systematic pattern recognition, rule application, and structured explanation generation align perfectly with machine capabilities while addressing genuine learner needs.

The even distribution across all skill challenges (vocabulary 11.7%, reading 10.5%, pronunciation 9.2%) suggests learners face interconnected difficulties rather than isolated problems. This validates our translation-based approach, which naturally integrates grammar learning with vocabulary expansion and contextual understanding.

User Profile and Readiness

Our target users emerge as a technologically experienced cohort: 75.7% have AI experience and appear to expect adaptive responses based on their technology adoption patterns. The correlation data reveals younger learners with higher English proficiency (age-proficiency: -0.46) and confidence (age-confidence: -0.39), potentially representing a generation benefiting from improved educational standards who seek refinement rather than remediation.

The "intermediate plateau" phenomenon, with most learners concentrated at Level 3 confidence, identifies our optimal target: users with sufficient foundation to understand grammatical concepts but still requiring guidance for advancement. This positioning avoids both the frustration of overly basic tools and the complexity barrier of advanced linguistic analysis.

Core Design Principles

Three fundamental principles emerge from the data:

1. **Explanation Over Correction**: With 73.4% rating WHY understanding as highly important and a 2:1 preference for examples over answers, the system must prioritise explanatory depth. This appears to represent a learning preference for comprehension over memorisation, though whether this represents a shift from previous approaches requires longitudinal data to confirm.

2. **Confidence Through Understanding**: The strong proficiency-confidence correlation (0.81) validates building genuine competence rather than providing quick fixes. Each interaction should enhance both skill and self-assurance through clear, accessible explanations that demystify grammatical complexity.

3. **Accessibility and Flexibility**: Feature requests clustering around mobile design and audio support reflect practical learning constraints. Professionals and students require tools accessible during commutes, breaks, or brief study windows rather than extended dedicated sessions. The system must support interrupted learning patterns, allowing users to engage meaningfully even in short time segments.

Implementation Priorities

The uniform mediocrity of current learning methods (3.2-3.5 effectiveness ratings) indicates market opportunity rather than user apathy. Learners actively seek better solutions, with the correlation between WHY importance and tool adoption likelihood (0.35) identifying our early adopters: engaged learners who value understanding.

The universal appeal of explanatory features across all demographic segments (correlations <0.35 with age, proficiency, and background) confirms that explanation-focused functionality should be the core value proposition, not an optional enhancement. Alternative translations ranking highly suggests learners seek exploration of possibilities rather than singular corrections. The system should reveal language richness rather than enforce rigid rules.

These implications directly shape the technical requirements and architectural decisions presented in subsequent sections, ensuring our system design remains grounded in empirical user needs rather than theoretical assumptions.

3.3 Technical Approaches Analysis

After establishing user needs through empirical research, this section analyses the technical approaches available to address the identified challenges. This analysis directly informs the requirements and design decisions presented in subsequent sections.

3.3.1 System Architecture Alternatives

Multiple technical approaches could potentially address the identified challenges of providing integrated translation and grammar explanation for Thai English learners.

Table 3.1: System Architecture Alternatives Analysis

Architecture	**Advantages**	**Limitations**
Monolithic End-to-End	• Reduced error propagation between components • Potentially faster inference • Simplified deployment • Unified optimisation across all tasks	• Limited interpretability for educational applications • Difficulty debugging specific component failures • Substantial training data requirements for multi-task objectives • Challenges in updating individual capabilities
Modular Pipeline	• Component-level optimisation • Interpretable intermediate outputs • Independent component updates • Educational transparency in processing stages	• Potential error propagation • Increased system complexity • Longer processing times • Coordination challenges between components

Based on user study findings showing 73.4% rate understanding WHY specific grammar choices as high importance, the modular approach provides the transparency necessary for effective educational delivery.

3.3.2 Component Implementation Alternatives

Table 3.2: Translation Component Approach Analysis

Approach Category	**Advantages**	**Limitations**	**Assessment**
Commercial API Services	• Superior translation quality • Reliable infrastructure • Automatic updates • No model maintenance	• External dependencies • Ongoing operational costs • Limited customisation • Privacy concerns	High quality, dependency trade-offs
Open-source Multilingual Models	• Full local control • No ongoing costs • Customisable for education • Research transparency	• Extensive fine-tuning required • Variable quality outcomes • High computational requirements • Technical expertise needed	Good control, considerable effort
Language-pair Specialised Models	• Optimised for Thai-English • Balanced performance-cost • Educational context awareness • Cultural nuance handling	• Limited language coverage • Model availability constraints • Version dependency risks • Specialised knowledge required	Targeted optimisation, limited scope

Table 3.3: Grammatical Analysis Approach Analysis

Approach Category	**Advantages**	**Limitations**	**Assessment**
Rule-based Linguistic Systems	• Fully interpretable logic • No training data requirements • Deterministic behavior • Expert knowledge integration	• Limited complex construction coverage • Difficulty with edge cases • Extensive linguistic expertise required • Maintenance overhead	High interpretability, limited coverage
Statistical Machine Learning	• Data-driven pattern recognition • Handles complex feature interactions • Adaptable to new domains • Quantifiable performance metrics	• Requires substantial training data • Limited interpretability • Domain-specific optimisation needed • Feature engineering complexity	Good performance, requires data
Neural Language Models	• Contextual understanding • Cross-lingual capabilities • Resilient to input variations • Strong multilingual performance	• Requires extensive training data • High computational requirements • Limited interpretability • Complex architecture management	High capability, resource intensive

Table 3.4: Educational Content Generation Approach Analysis

Approach Category	**Advantages**	**Limitations**	**Assessment**
Template-based Systems	• Consistent pedagogical structure • Predictable output quality • Educational content control • No inference costs	• Limited contextual flexibility • Potential repetitiveness • Extensive template development • Difficulty handling edge cases	Consistent but inflexible
Generative AI Systems	• Contextual adaptability • Natural language fluency • Creative explanation approaches • Multilingual capabilities	• Variable output quality • Potential factual inaccuracies • Limited pedagogical structure control • Computational requirements	Flexible but unpredictable
Hybrid Rule-Generation Systems	• Structured pedagogical framework • Contextual adaptation • Quality control mechanisms • Educational objective alignment	• Development complexity • Integration challenges • Maintenance overhead • Technical expertise required	Balanced approach, complex implementation

3.3.3 Deployment Architecture Considerations

Table 3.5: Deployment Architecture Alternatives Analysis

Approach	**Advantages**	**Limitations**	**Assessment**
Local Deployment	• Complete control • No external dependencies • Predictable costs • Data privacy	• High computational requirements • Infrastructure complexity • Limited scalability • Maintenance burden	High control, resource intensive
Cloud API Integration	• Reduced infrastructure complexity • Access to high-quality models • Automatic scaling • Lower entry barriers	• External dependencies • Variable costs • Potential privacy concerns • Service availability risks	Low complexity, dependency trade-offs
Hybrid Architecture	• Optimised component placement • Cost-performance balance • Risk distribution • Flexibility in optimisation	• Increased architectural complexity • Multiple integration points • Coordinated failure management • Development overhead	Optimal balance, higher complexity

3.3.4 Technical Approach Selection Rationale

Based on the user study findings and technical analysis, the optimal approach combines:

- **Modular Pipeline Architecture**: Provides the transparency necessary for educational understanding, aligning with the 73.4% of users who value WHY explanations
- **Specialised Thai-English Models**: Addresses the specific linguistic challenges faced by our target demographic, where grammar ranks as the third most common challenge (13.4%)
- **Neural Classification with Hierarchical Structure**: Delivers the accuracy required while maintaining interpretability for learners at the intermediate plateau (Level 3 confidence)
- **Generative AI with Template Post-processing**: Combines the flexibility needed for diverse grammatical contexts with the consistency required for reliable educational content
- **Hybrid Deployment**: Balances computational efficiency with cost-effectiveness, crucial for sustained educational accessibility

This technical foundation directly informs the system requirements and design decisions presented in the following sections.

3.4 Requirements Analysis

Based on the user study findings and technical analysis, system requirements that address both technical capabilities and pedagogical effectiveness are derived.

3.4.1 Functional Requirements

Table 3.6: Functional Requirements Specification

ID	**Requirement Description**	**Priority**
FR1	The system must accurately translate Thai sentences to grammatically correct English with >85% semantic accuracy	High
FR2	The system must identify and classify the primary tense used in the English translation with >80% accuracy	High
FR3	The system must generate educational explanations for why specific tenses are used in given contexts	High
FR4	The system must detect incomplete sentences and provide educational feedback rather than processing fragments	Medium
FR5	The system must provide contextual examples demonstrating correct tense usage	Medium
FR6	Explanations must address specific challenges of Thai learners (absence of tense markers in L1)	High
FR7	Grammar explanations must be accessible to intermediate learners (B1-B2 proficiency levels)	High
FR8	The interface must support both Thai and English languages for user accessibility	Medium
FR9	The system must allow users to rate translation and explanation quality	Low
FR10	The system must store user feedback for continuous improvement analysis	Low

3.4.2 Technical Architecture Requirements

Table 3.7: Technical Architecture Requirements

ID	Requirement Description	Priority
TR1	Modular pipeline architecture with separate components for translation, classification, and explanation generation	High
TR2	Thai-English specialised translation models trained on domain-specific corpora for cultural and linguistic accuracy	High
TR3	Hierarchical tense classification system supporting both coarse-grained (3) and fine-grained (24) categories	High
TR4	Educational NLP capabilities for generating explanations that address Thai L1 interference patterns	High
TR5	Hybrid deployment architecture combining local models (translation/classification) with API services (explanation)	Medium
TR6	Fragment detection system to handle incomplete input appropriately	Medium
TR7	Cloud-based infrastructure supporting horizontal scaling and high availability	Medium
TR8	API integration capabilities for external language model services	High
TR9	Data collection and analytics infrastructure for continuous system improvement	Low

Chapter 4

Design and Implementation - Core Models and Functionality

This chapter presents the design and implementation of the three core models that enable Thaislate's educational capabilities: a translation language model, a tense classification model, and an explanation language model. Each model was selected and optimised to address specific challenges identified through the user study: accurate Thai-English translation, precise tense classification, and pedagogically effective explanation generation. Rather than attempting to use a single general-purpose language model for all tasks, specialised components were developed that excel at their individual functions.

The decision to use three distinct models emerged from early experimentation revealing that instruction-following language models, while capable of handling multiple tasks, frequently failed at accurate tense classification, a critical flaw that would undermine the entire educational purpose. By separating concerns and optimising each model for its specific function, significantly better performance was achieved than monolithic approaches, with individual models demonstrating 93.2% translation fluency, 94.7% tense classification accuracy, and 84.9% explanation quality in isolated evaluation.

4.1 Model Selection and Design Philosophy

4.1.1 Core Educational Challenges and Model Requirements

The user study revealed specific technical requirements that directly influenced the model selection strategy:

1. **Translation Accuracy for Learning**: Translation models must preserve temporal markers from Thai that inform English tense selection, rather than simply producing fluent output

2. **Specialised Classification Need**: Testing demonstrated that general-purpose language models frequently misclassified tenses (leading to incorrect explanations), necessitating a dedicated classification component

3. **Pedagogical Explanation Generation**: Users prioritised understanding WHY specific tenses are used, requiring explanation models trained for educational contexts rather than general conversation

These requirements established three core model design principles:

1. **Specialised Excellence**: Each model should excel at its specific educational task rather than attempting multiple functions
2. **Educational Optimisation**: Models must be optimised for learning outcomes, not just technical performance metrics
3. **Practical Deployment**: Model selection must balance ideal functionality with real-world computational constraints

4.2 Development Journey and Architectural Evolution

The development of Thaislate evolved through three distinct phases that fundamentally shaped the final system architecture. Rather than following a predetermined technical plan, the project's development represents a journey of discovery where initial limitations revealed the need for specialised approaches, ultimately leading to innovative solutions for educational natural language processing.

4.2.1 Phase 1: Initial Vision and Critical Discovery

The project began with a straightforward goal: create "an English grammar breakdown thing" using a single, detailed language model. This approach initially seemed logical. Instruction-following models had demonstrated capabilities in educational content generation and could theoretically handle both translation and grammar explanation tasks within a unified framework.

During early experimentation, a critical limitation emerged that would reshape the entire project architecture. Instruction-following models, while capable of generating fluent and pedagogically appropriate explanations, consistently misclassified English tenses. This created a devastating problem for educational applications: learners would receive confident, well-structured explanations about incorrect grammatical concepts.

This discovery necessitated a complete architectural redesign. The realisation that explanation quality meant nothing if built upon incorrect classification forced the development toward specialised components, a decision that would prove central to the project's eventual success.

4.2.2 Phase 2: Pipeline Architecture Discovery

The three-stage pipeline architecture emerged organically through trial-and-error within Google Colab's experimental environment, rather than through systematic design. While testing different model combinations and approaches, it became evident that specialised models working in sequence could achieve superior results compared to any single general-purpose system.

The pipeline architecture was not planned but discovered during experimental work. Testing various model combinations revealed that translation → classification → explanation as separate, specialised processes outperformed all attempts at unified models. This discovery became the foundation of the entire system design.

4.2.3 Phase 3: Production Migration and Practical Constraints

The transition from Colab's experimental environment to production deployment revealed numerous practical challenges that shaped the final system architecture.

Colab's inability to support real user testing necessitated a complete platform migration to Google Cloud. This transition required learning deployment technologies, developing web interfaces, and managing production constraints that had not been considered during the experimental phase.

Memory and computational limitations led to several solutions that ultimately improved the system:

- **GGUF (GPT-Generated Unified Format) Quantisation Discovery**: Initial attempts to load multiple models exceeded 16GB RAM limits, leading to the adoption of GGUF quantisation that enabled efficient CPU-only inference
- **Hybrid Architecture Evolution**: Resource constraints drove the hybrid local-API approach, combining local efficiency with cloud-based quality for optimal cost-performance balance
- **Thread-Safe Design**: Concurrent user access issues necessitated thread-safe model management, improving system reliability beyond initial requirements

Throughout this phase, educational effectiveness consistently took precedence over technical elegance. Fragment detection, graceful error handling, and user-friendly messaging were prioritised based on real learner needs rather than computational optimisation.

4.2.4 Three-Model Architecture Overview

The development journey culminated in a three-model architecture where each component serves a distinct educational function, validated through practical deployment and user testing:

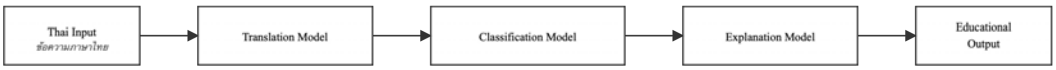


Figure 4.1: Core Model Architecture for Educational Language Processing

This approach provides several key advantages over monolithic models:

- **Task-Specific Optimisation**: Each model can be fine-tuned for its particular educational function without compotential others
- **Performance Reliability**: Each model demonstrates higher accuracy than general-purpose alternatives in comparative evaluation
- **Independent Improvement**: Models can be upgraded or replaced without affecting the entire system
- **Computational Efficiency**: Resource allocation can be optimised for each model's specific requirements

This chapter examines each of the three core models in detail, focusing on their individual design decisions, implementation approaches, and performance characteristics. Chapter 5 will demonstrate how these components integrate into a complete educational system.

4.3 Translation Model: Typhoon Translate 4B

The translation component represents the first critical step in transforming Thai learner input into analysable English text for grammatical instruction. While the system uses Typhoon Translate 4B with its standard prompting approach as recommended by the model developers, the model's Thai-English specialisation naturally preserves temporal markers and semantic relationships that are essential for subsequent tense analysis.

4.3.1 Thai-English Translation Challenges for Grammar Learning

Thai-English translation for grammar learning presents unique challenges that influenced the model selection:

1. **Temporal Marker Preservation**: Thai expresses time relationships through context words (meuua waan nii for "yesterday", ja for future intentions) rather than verb conjugation. The translation model must recognise these markers and convert them appropriately to English tense structures without losing the temporal information needed for subsequent classification.

2. **Semantic Accuracy Over Fluency**: For educational purposes, maintaining the exact meaning intended by the learner takes precedence over producing idiomatic English, as misinterpretations could lead to incorrect grammar explanations.

4.3.2 Typhoon Translate Model Selection and Optimisation

The choice of Typhoon Translate 4B emerged from systematic evaluation of Thai-English translation models for educational applications:

Typhoon models are specifically trained on Thai-English parallel data, unlike general multilingual models that treat Thai as one of hundreds of languages. This specialisation proved crucial for handling Thai temporal markers and cultural context that appear in learner input. Comparative testing between Typhoon Translate 4B and instructing Typhoon 2.1 12B to perform translation showed that the dedicated translation model was both faster and produced more consistent results for educational contexts.

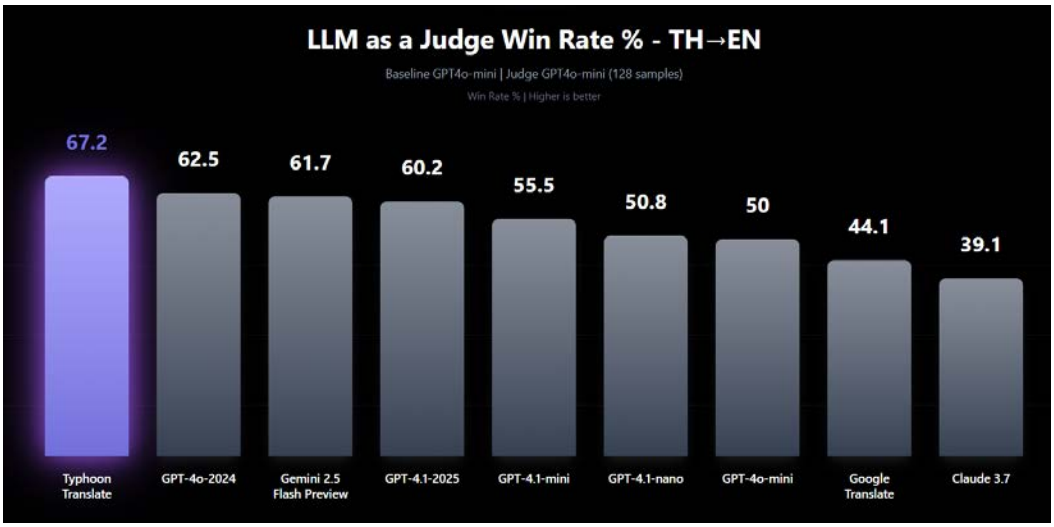


Figure 4.2: Typhoon Translate Performance Comparison for Thai-English Translation (Source: Typhoon AI)

Figure 4.2 demonstrates Typhoon Translate's superior performance in Thai-English translation, achieving a 67.2% win rate against major competitors including GPT-4 variants, Claude 3.7, and Google Translate (44.1%).

Converting to GGUF (GPT-Generated Unified Format) quantisation enabled deployment on CPU-only infrastructure, crucial for cost-effective deployment. This quantisation approach provides substantial computational advantages: faster inference times, reduced memory requirements, and elimination of GPU dependency, enabling deployment on standard cloud instances.

4.3.3 Prompting Strategy for Educational Translation

Based on guidance from the Typhoon model developers[1], the system employs a straightforward prompting approach optimised for educational contexts. The developers recommended using simple, direct prompts rather than complex instruction formatting to achieve optimal translation quality for Thai-English language pairs.

The translation prompt used in the pipeline is:

```
Translate this Thai text to English: {thai_input}
```

This minimalist approach was specifically recommended by the model creators and reflects the specialised nature of Typhoon Translate, which was trained specifically for Thai-English translation tasks and achieves optimal performance through simple prompts without requiring complex instructional formatting or risking prompt-induced errors that could affect educational accuracy.

4.4 Tense Classification Model: Custom XLMRoberta Model

The tense classification component represents the core educational innovation of Thaislate, implementing a complex 24-category hierarchical classification system specifically designed for Thai English learners.

4.4.1 The Critical Need for Specialised Tense Classification

The development of a dedicated tense classification model emerged from a fundamental limitation discovered during system testing: general-purpose language models consistently failed at accurate tense identification, which would have undermined the entire educational purpose of the system.

Initial experiments with GPT-style instruction-following models seemed potentially viable as they could handle translation and explanation generation reasonably well. However, when tasked with tense classification, these models demonstrated concerning inconsistency. A sentence like "I have been studying for three years" might be classified as "Present Continuous" on one attempt and "Present Perfect Continuous" on another, despite being identical input. This unpredictability would be pedagogically harmful, as learners could receive conflicting explanations for the same grammatical structure.

Unlike general NLP applications where small classification errors might be acceptable, educational systems require high reliability. An incorrect tense classification leads directly to incorrect grammar explanation, potentially reinforcing learner errors rather than correcting them. For Thai learners already struggling with English temporal concepts, such misinformation could significantly impede learning progress.

This discovery necessitated developing a classification system optimised specifically for educational reliability:

- **Consistency**: Identical inputs must produce identical classifications every time

[1] https://huggingface.co/scb10x/typhoon-translate-4b

- **Nuanced Distinction**: Ability to distinguish between subtle tense variations crucial for Thai learners (e.g., habits vs. general facts)
- **Confidence Scoring**: Quantified certainty levels to identify potentially problematic classifications
- **Educational Taxonomy**: Classification categories aligned with how tenses are taught to Thai learners

4.4.2 The 24-Category Tense Taxonomy

The classification system implements a custom taxonomy developed with reference to established Thai grammar teaching materials [16], designed to address specific challenges Thai learners face:

Code	Thai Category	Example Usage
Present Tenses (9 categories)		
HABIT	Habit/Routine	She drinks tea every morning
FACT	General Fact	The sun rises in the east
SCHEDULEDFUTURE	Scheduled/Planned	The train departs at 6:45
SAYING	Proverb/Saying	Actions speak louder than words
HEADLINE	News Headline	Mayor announces new plan
HAPPENING	Currently Happening	They are discussing the budget
NOWADAYS	Present Continuous	Trains are getting more expensive nowadays
SUREFUT	Planned Future	I am meeting the client at noon
PROGRESS	Ongoing Change	The crew is building the bridge
Perfect Tenses (4 categories)		
JUSTFIN	Just Finished	She has just finished homework
RESULT	Present Result	I have lost my keys
EXP	Life Experience	I have visited Japan three times
SINCEFOR	Duration Period	He has been studying for two years
Past Tenses (5 categories)		
NORFIN	General Past	I did not eat the cake
INTERRUPT	Interrupted Action	I was watching TV when he called
DOINGATSOMETIMEPAST	Past Continuous	They were working at 8 PM
BEFOREPAST	Before Past Action	She had left before the party
DURATION	Past Duration	They had been driving for hours
Future Tenses (6 categories)		
50PERC	50% Prediction	It will rain soon
PROMISE	Promise/Commitment	I will always support you
RIGHTNOW	Future Simple	I'll open the window right now
LONGFUTURE	Distant Future	She will be studying in Paris
PREDICT	Future Prediction	Scientists will have developed a vaccine
WILLCONTINUEINFUTURE	Future Duration	I will have been working for hours

Table 4.1: Complete 24-Category Tense Classification Taxonomy

4.4.3 Dataset Preparation and Training

Dataset Sources and Composition

The classification model was trained on a carefully curated dataset combining multiple sources:

- 2,505 sentences with fine-grained labels (manually annotated by the author)
- 13,316 sentences with coarse labels only (Past/Present/Future)
- Total: 15,755 unique English sentences
- 80/20 train/validation split with stratification

Two Kaggle datasets [17, 18] were selected as base sources for their detailed coverage of English tense structures and educational appropriateness for language learning contexts.

```
🔎 Dataset summary
• fine-grained rows : 2505
• coarse-only rows  : 13316
• merged unique rows: 15755
```

	text	coarse_label	fine_label
8251	online forms will be designed with accessibili...	Future	None
1385	They hosted a picnic in the park last summer.	Past	NORFIN
15344	genetic engineering will influence avatars	Future	None
11499	voicecontrolled assistants will assist with ho...	Future	None
10412	restaurants are adapting their menus to accomm...	Present	None

Figure 4.3: Dataset Structure Example showing Hierarchical Labeling Approach with Fine-grained and Coarse Labels

Figure 4.3 illustrates the hierarchical structure of the training dataset, showing how sentences are labeled with both coarse temporal categories (Past/Present/Future) and fine-grained tense classifications. The dual-labelling approach enables the model to learn both broad temporal understanding and specific grammatical distinctions, with some sentences having only coarse labels to support the hierarchical training methodology.

Manual Annotation Process

The 2,505 fine-grained labels were manually annotated by analysing each sentence for its primary tense usage pattern and labelling according to the 24-category taxonomy framework. Labels were assigned based on Thai teaching methodology and learning objectives, with ambiguous cases resolved using context and primary grammatical intent. All annotations were thoroughly checked for consistency and accuracy by the author alone.

Ethical Considerations for Dataset Annotation

The dataset annotation process received separate ethical approval from the University Research Ethics Committee of the University of Sheffield (Application Number 070793). While the annotation involved only the author working with publicly available English sentences from Kaggle datasets, formal ethical review ensured appropriate research standards were maintained. The sentences contained no personally identifiable information and were already in the public domain for educational purposes. The annotation focused solely on grammatical classification for educational tool development. The complete ethics application and approval documentation for the dataset annotation work can be found in Appendix A.2.

4.4.4 Hierarchical Classification Design and Architecture

The core innovation of the tense classification system lies in its hierarchical approach, which mirrors how humans naturally understand temporal relationships by first identifying the broad time frame (past, present, future) before determining specific grammatical functions.

Traditional flat classification approaches treat all 12 tense categories as equally distinct, ignoring the natural relationships between them. The hierarchical approach adopted here recognises that "Past Simple" and "Past Continuous" are both types of past tenses, sharing temporal characteristics while differing in aspectual details. This structure provides three key benefits:

- **Pedagogical Alignment**: Matches how tenses are taught with broad concepts first, then specific applications
- **Error Mitigation**: If fine-grained classification fails, coarse classification can still provide useful educational feedback
- **Training Efficiency**: Shared temporal knowledge improves learning across related categories

The model employs a dual-head architecture where XLM-RoBERTa's multilingual encoder feeds into two classification heads:

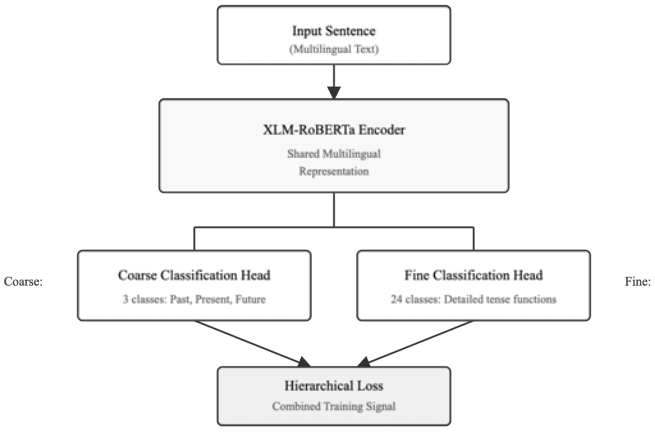


Figure 4.4: Hierarchical Tense Classification Architecture with XLM-RoBERTa Encoder

- **Coarse Head**: Predicts broad temporal categories (Past/Present/Future)
- **Fine Head**: Predicts specific tense functions (24 detailed categories)
- **Shared Encoder**: XLM-RoBERTa processes multilingual input understanding

The model uses weighted loss optimisation that balances both classification levels:

- **70% weight on fine-grained accuracy**: prioritises the detailed classifications needed for grammar instruction
- **30% weight on coarse accuracy**: Ensures temporal consistency and provides fallback capability
- **Combined loss function**: Prevents the model from achieving fine-grained accuracy at the expense of basic temporal understanding

This hierarchical approach provides theoretical advantages for both training efficiency and educational application.

4.5 Grammar Explanation Model: Typhoon 2.1 12B

The grammar explanation component represents the final and most complex challenge in educational language processing systems: transforming technical tense classifications into clear, helpful explanations that actually improve Thai learners' understanding of English grammar. Unlike translation or classification, which have objective correctness measures, explanation quality depends on pedagogical effectiveness, a significantly more nuanced challenge requiring specialised model selection and optimisation.

4.5.1 Pedagogical Design Based on User Study

The explanation structure was directly informed by user study findings (Chapter 3), where learners expressed specific needs for understanding English grammar:

The three core components address specific user needs identified in the survey:

1. **Tense Analysis**: Clear explanation of how the tense is formed grammatically and its usage context, addressing the primary grammar learning challenge
2. **Vocabulary Focus** (22.1% struggled with vocabulary): Key terms and translation choices from the analysed sentence
3. **Common Mistakes** (Error prevention focus): Addresses typical errors Thai learners make with this tense

This three-component structure ensures detailed support addressing both primary grammatical understanding and secondary vocabulary learning challenges identified through user research.

4.5.2 Model Selection and API Integration

The choice of Typhoon 2.1 12B via Together.ai API[2] represents a strategic decision prioritising educational quality over computational independence. Unlike the translation and classification components which operate locally, the explanation model requires complex language generation capabilities that justify API-based deployment.

Model Architecture Rationale: Typhoon 2.1 12B provides the linguistic sophistication necessary for generating pedagogically appropriate explanations tailored to Thai learners. The larger parameter count enables nuanced understanding of cross-linguistic challenges and supports consistent adherence to structured explanation templates.

API Integration Decision: The API-based approach offers several architectural advantages over local deployment:

- **Quality prioritisation**: Educational effectiveness takes precedence over computational independence
- **Infrastructure efficiency**: Eliminates need for hosting additional large models locally
- **Scalability**: API infrastructure handles variable load without local resource management
- **Model updates**: Automatic access to model improvements without local redeployment

4.5.3 Dual Prompting Strategy Architecture

The explanation model employs two distinct prompting strategies based on input classification: one optimised for complete sentences requiring full tense analysis, and another designed for sentence fragments that need educational guidance without grammatical analysis. This dual approach ensures appropriate educational responses regardless of input completeness.

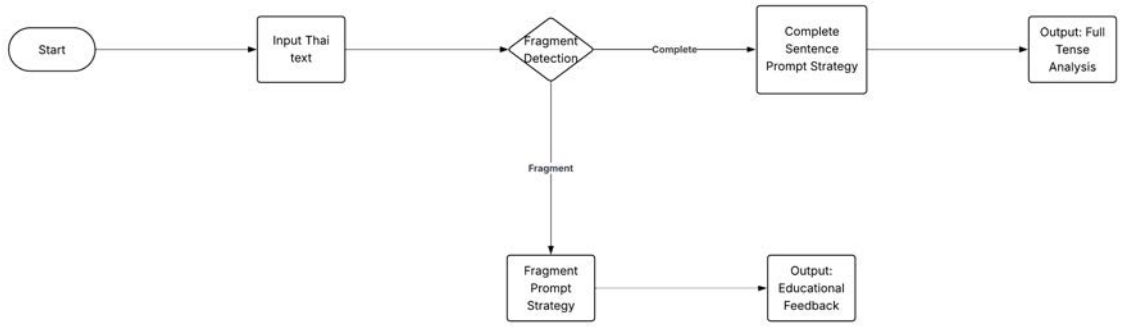


Figure 4.5: Dual Prompting Strategy Decision Flow: Complete Sentence vs Fragment Processing

Input-Dependent Strategy Selection: The system determines prompting strategy through fragment detection logic implemented in the preprocessing pipeline. Complete

[2]https://www.together.ai/models/typhoon2-1-gemma3-12b

sentences proceed through the full translation-classification-explanation pipeline, while fragments bypass tense classification entirely and receive a different educational feedback about sentence formation and completion.

Strategy 1: Complete Sentence Processing: For grammatically complete inputs, the system employs detailed Thai-language prompts that incorporate:

- **Pedagogical Context Setting**: Establishes AI role as English grammar teacher for Thai students
- **Tense-Specific Knowledge Integration**: Incorporates detailed information about the classified tense from the 24-category taxonomy
- **Structured Response Requirements**: Mandates three-component output (tense analysis, vocabulary, common mistakes)
- **Cross-Linguistic Explanation**: Requires explanations that bridge Thai-English grammatical differences

Strategy 2: Fragment Educational Response: For incomplete inputs identified as fragments, the system uses alternative prompting that focuses on:

- **Educational Explanation**: Clarifies why tense analysis cannot be performed on incomplete sentences
- **Meaning Clarification**: Explains the meaning and potential usage of the provided text fragment
- **Sentence Completion Guidance**: Provides specific suggestions for forming grammatically complete sentences

Table 4.2: Dual Prompting Strategy Comparison

Aspect	**Complete Sentence Strategy**	**Fragment Strategy**
Input Type	Grammatically complete sentences with identifiable verbs	Incomplete sentences, single words, or phrases
Classification	Full tense classification using XLM-RoBERTa model	Bypasses tense classification entirely
Response Focus	Detailed grammatical analysis and tense explanation	Educational guidance on sentence formation
Output Structure	Three-component format (tense, vocabulary, mistakes)	Educational explanation with completion guidance
Educational Goal	Deepen understanding of specific tense usage	Teach sentence structure and completeness concepts
Language	Primarily Thai with English examples	Thai language with structural guidance

4.5.4 Confidence-Aware Response System

The explanation model incorporates confidence information from the XLM-RoBERTa classifier to provide more honest and educationally appropriate responses when tense classification is uncertain.

Confidence Integration Through Prompting: The confidence score (derived from the XLM-RoBERTa model's softmax probabilities) is included directly in the prompt sent to the Typhoon 2.1 12B explanation model. The prompt contains the confidence percentage and specific instructions that guide the model to adjust its response style based on three confidence tiers:

- **High Confidence (>90%)**: The prompt instructs the model to provide clear, definitive explanations with confident language
- **Medium Confidence (70-90%)**: The prompt includes instructions to mention that the system is moderately confident and suggest alternative tense possibilities that might be relevant
- **Low Confidence (<70%)**: The prompt explicitly instructs the model to use hedging language (Thai equivalents of "might be", "probably", "there's a possibility") while providing alternative interpretations

Educational Benefits of Confidence Integration: This prompt-based approach provides key educational advantages: it ensures the explanation model acknowledges uncertainty when the classifier is less confident, prevents overconfident explanations of potentially incorrect classifications, and teaches students that grammar analysis can involve multiple valid interpretations. By including confidence information in the prompt, the system maintains educational honesty while still providing useful learning content even when tense classification is uncertain.

4.5.5 Response Formatting and Output Processing

The explanation model generates unstructured text that requires complex processing to create the organised, educational content presented to users. This output parsing system transforms raw API responses into structured sections that facilitate effective grammar learning.

Multi-Stage Processing Pipeline

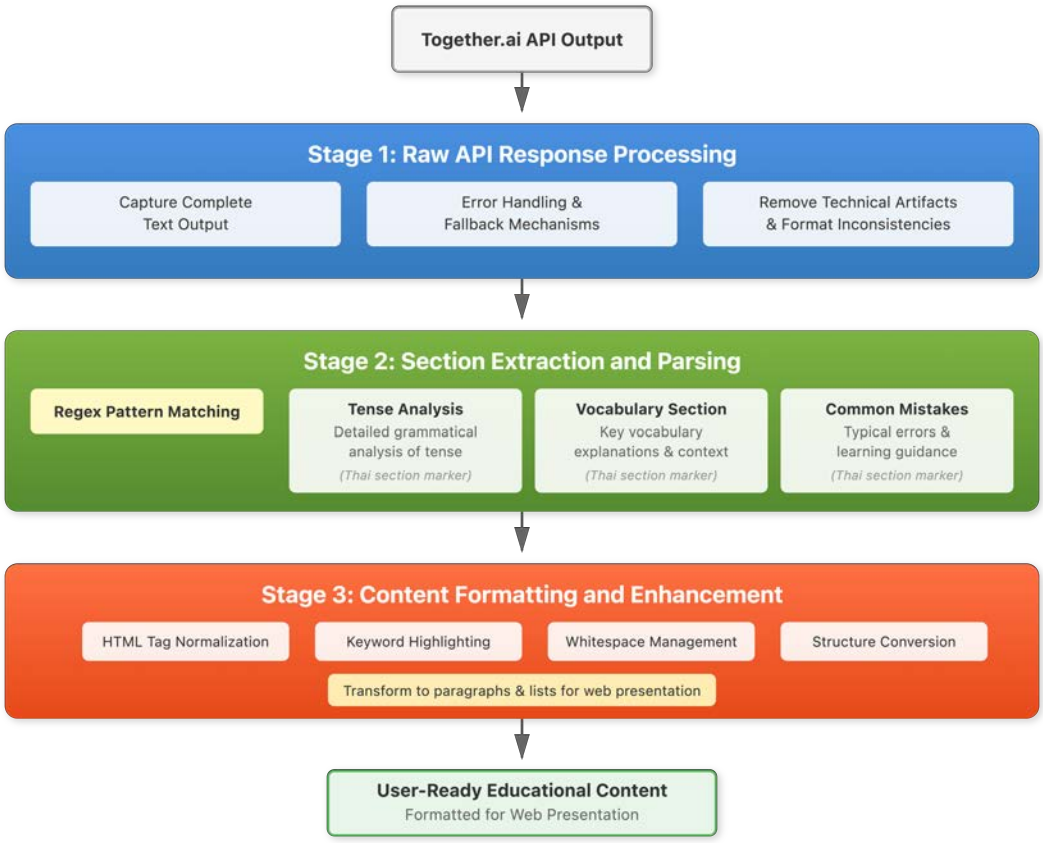


Figure 4.6: Three-stage processing pipeline for transforming raw LLM output into structured educational content

The output processing system transforms raw API responses into structured educational content through the three-stage pipeline shown in Figure 4.6. Each stage progressively refines the content:

Stage 1: Raw API Response Processing - Captures the Together.ai API output and handles potential errors through fallback mechanisms while removing technical artifacts.

Stage 2: Section Extraction and Parsing - Uses regex pattern matching to identify Thai section markers and extract the three core educational components (Tense Analysis, Vocabulary, and Common Mistakes).

Stage 3: Content Formatting and Enhancement - Applies HTML normalisation, keyword highlighting, and structure conversion to create web-ready educational content.

Adaptive Content Structuring

The parsing system adapts its output format based on input classification results, providing different content structures for different learning scenarios:

Complete Sentence Processing: Standard three-section format with detailed tense analysis, relevant vocabulary explanations, and common mistake patterns that help learners understand both the specific usage and broader learning context.

Fragment Input Handling: Modified section titles and content focus that explain why grammatical analysis cannot proceed, provide relevant vocabulary guidance, and offer suggestions for forming complete sentences suitable for tense analysis.

Error Recovery and Fallback Systems

The system implements basic error handling to ensure users receive some educational content even when API calls fail:

- **Mock Explanation Fallback**: When the Together.ai API is unavailable or fails, the system returns pre-written mock explanations based on the coarse tense classification (Present/Past/Future)
- **Failed Section Extraction**: When regex parsing cannot extract expected sections from the API response, default text is provided indicating the section could not be parsed
- **Confidence-Aware Fallbacks**: Mock explanations adjust their language based on classification confidence, adding hedging phrases for low-confidence predictions

While these fallback mechanisms ensure the system remains functional during API failures, they provide limited educational value compared to the full AI-generated explanations.

4.6 Core Models Summary

This chapter has presented the design and implementation of the three core models that enable Thaislate's educational functionality. Each model was carefully selected and optimised for specific educational challenges identified through user research and pedagogical requirements.

Translation Language Model (Typhoon Translate 4B) employs GGUF quantisation for cost-effective CPU-only deployment while preserving temporal markers from Thai input essential for downstream grammatical analysis. The Thai-English training enables handling of cultural context and temporal expressions that general multilingual models struggle with.

Tense Classification Model (Custom XLM-RoBERTa) represents the core educational innovation through its hierarchical architecture and 24-category taxonomy designed specifically for Thai learners. The model's multilingual foundation supports cross-linguistic understanding while the dual-head design ensures both broad temporal consistency and fine-grained pedagogical precision.

Explanation Language Model (Typhoon 2.1 12B) transforms technical classifications into pedagogically structured content through complex Thai-language prompting and API-based generation. The architectural decision to prioritise educational quality over computational independence demonstrates how learning effectiveness requirements can justify external dependencies.

Chapter 5

Design and Implementation - Pipeline and Website

This chapter presents the integration of the three specialised models from Chapter 4 into a cohesive educational system, focusing on pipeline orchestration and web application development, and addresses the preprocessing systems that prepare user input for model processing, the coordination mechanisms that manage component interactions, the web application features that support educational effectiveness, and the production deployment strategy that enables real-world user testing and validation. The implementation, built using Flask 2.3.3 web framework, leverages multiple specialised libraries: llama-cpp-python for GGUF model inference (Typhoon Translate), Transformers library with PyTorch for the XLM-RoBERTa classifier, and the Together API client for explanation generation. The system also incorporates Flask-SQLAlchemy 3.0.5 for user management, Flask-Login 0.6.3 for authentication, and Flask-Babel 2.0.0 for bilingual interface support.

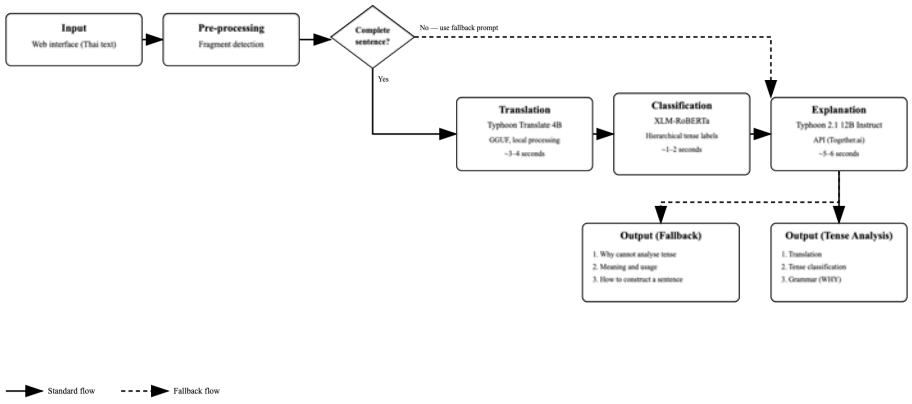


Figure 5.1: Complete Thaislate System Architecture

5.1 Pipeline Architecture Overview

5.1.1 Model Interaction and Data Flow

The pipeline orchestrates three specialised models in sequential processing to transform Thai learner input into detailed English grammar instruction. Each model performs a distinct function within the educational workflow, with outputs from one stage providing inputs for the next.

The integration strategy emphasises data flow optimisation and educational coherence. The translation stage produces English text optimised for grammatical analysis rather than natural fluency. The classification stage processes this structured English to identify temporal patterns and tense categories. The explanation stage synthesises the classification results with the original Thai context to generate pedagogically appropriate content for Thai learners.

5.1.2 Component Integration Strategy

The pipeline architecture implements a modular integration strategy that coordinates the three models while maintaining fault tolerance and educational effectiveness. The sequential processing design allows each component to specialise in its educational function without compromising overall system performance.

The integration strategy addresses several technical challenges: data format consistency between model outputs and inputs, error propagation management across pipeline stages, and educational context preservation throughout the processing chain. The modular approach enables independent optimisation of each component while ensuring seamless data handoff between processing stages.

5.2 Preprocessing Pipeline

Before any model processing occurs, the system implements detailed preprocessing to ensure appropriate educational content and optimal learning conditions. These preprocessing systems emerged from practical deployment experience and user feedback, addressing real-world challenges not initially anticipated in the design phase.

Table 5.1: Preprocessing Pipeline Components and Functions

Component	**Function**
Educational Content Filtering	Detects inappropriate content using 40+ Thai profanity patterns with exception handling for legitimate words
Language Detection	Identifies English vs Thai input through character distribution analysis
Token Limitation	Restricts input to 100 tokens maximum for computational efficiency
Multi-Sentence Detection	Identifies sentence boundaries using regex patterns for Thai and English punctuation
Fragment Detection	Validates complete sentence structure including length requirements, verb presence, and structural patterns
Pre-processing Validation	Performs detailed input checks before any language model calls
Graceful Degradation	Enables partial processing with educational feedback when validation issues occur

5.2.1 Input Validation and Safety Systems

The profanity filtering system implements a dual-language approach designed specifically for educational contexts, recognising that inappropriate content not only violates platform guidelines but also interferes with effective learning outcomes.

The system employs pattern-based detection using over 40 Thai profanity patterns, including both direct matches and contextual variations that account for common spelling adaptations and spacing modifications. The detection system includes detailed exception patterns to prevent false positives with legitimate Thai words that might share similar character sequences, ensuring that educational discussions about sensitive topics such as relationships, health, and social issues are not inappropriately blocked while maintaining content standards.

When inappropriate content is detected, the system provides gentle educational guidance rather than harsh rejection, framing content filtering as optimisation for learning rather than censorship.

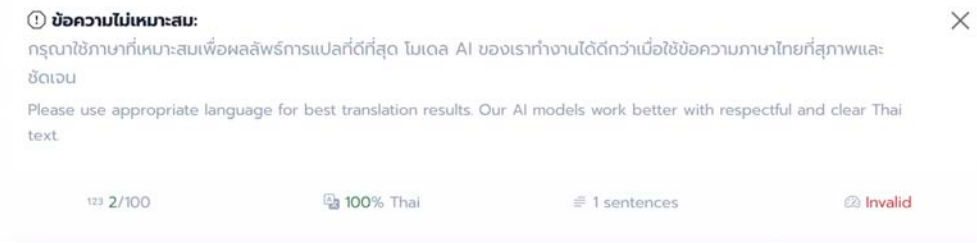


Figure 5.2: Educational Content Filtering Warning Interface

For language detection, the system addresses common user errors where learners input English text instead of Thai by implementing character distribution analysis and providing educational guidance for optimal system usage.

Figure 5.3: Language Detection Warning System

The token limitation system (100 tokens maximum) serves dual purposes: computational efficiency and educational effectiveness.

Single-sentence input produces more accurate tense classification and clearer grammatical explanations, as the classifier model was trained on simple sentences. Multi-sentence inputs often contain mixed tenses, leading to confusion in educational contexts.

The system implements Thai-specific token counting that accounts for the scriptural differences between Thai and English, recognising that Thai characters typically require approximately three characters per token compared to English word-based tokenisation.

When limits are exceeded, the system provides educational context rather than simple rejection, explaining that shorter, single-sentence inputs produce better learning results. This approach teaches users optimal usage patterns while maintaining system performance constraints.

The multi-sentence detection and processing system identifies sentence boundaries in inputs and provides appropriate user guidance while continuing to process the first sentence.

The system uses regular expression patterns to identify sentence boundaries based on common punctuation markers in both Thai and English text, counting meaningful sentence segments after filtering empty strings.

When multiple sentences are detected, users receive informative guidance explaining that the system will analyse only the first sentence for tense classification and explanation, providing transparency about system behaviour while maintaining educational value.

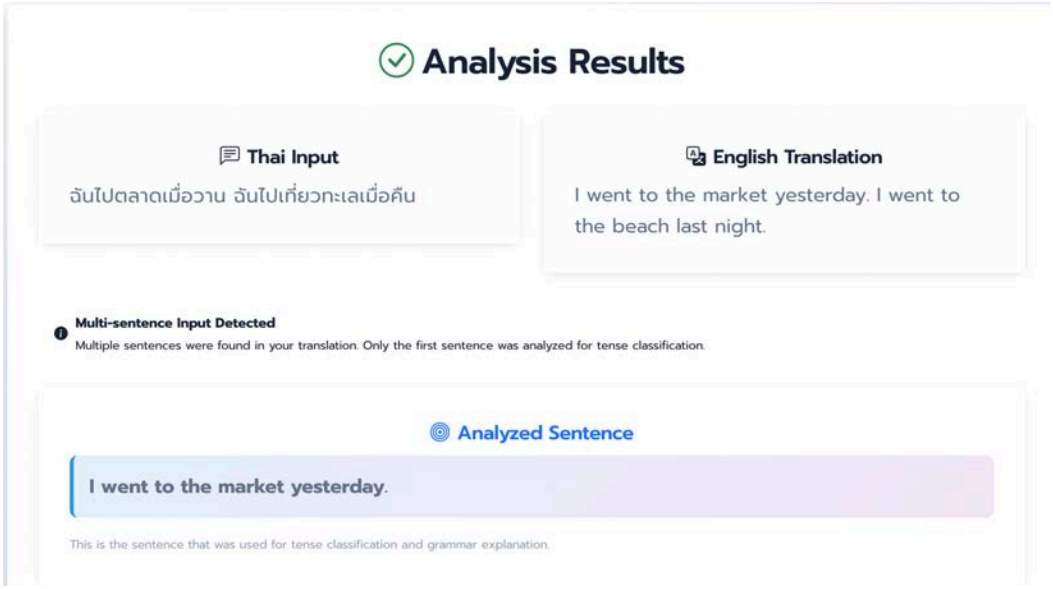


Figure 5.4: Multi-Sentence Detection and Processing Interface

5.2.2 Fragment Detection System

The system determines whether input constitutes a complete sentence suitable for tense classification before proceeding with grammatical analysis. This prevents inappropriate analysis of incomplete inputs while providing educational guidance.

The fragment detection methodology employs rule-based analysis examining three key criteria:

- Length requirements include minimum word and character counts with exceptions for pronoun-verb combinations
- Structural pattern recognition identifies common fragments like prepositional phrases and isolated words
- Verb presence validation checks against a detailed English verb database including auxiliary and action verbs

When fragments are detected, the system generates educational feedback explaining why tense analysis cannot proceed and provides guidance on forming complete sentences suitable for grammatical analysis.

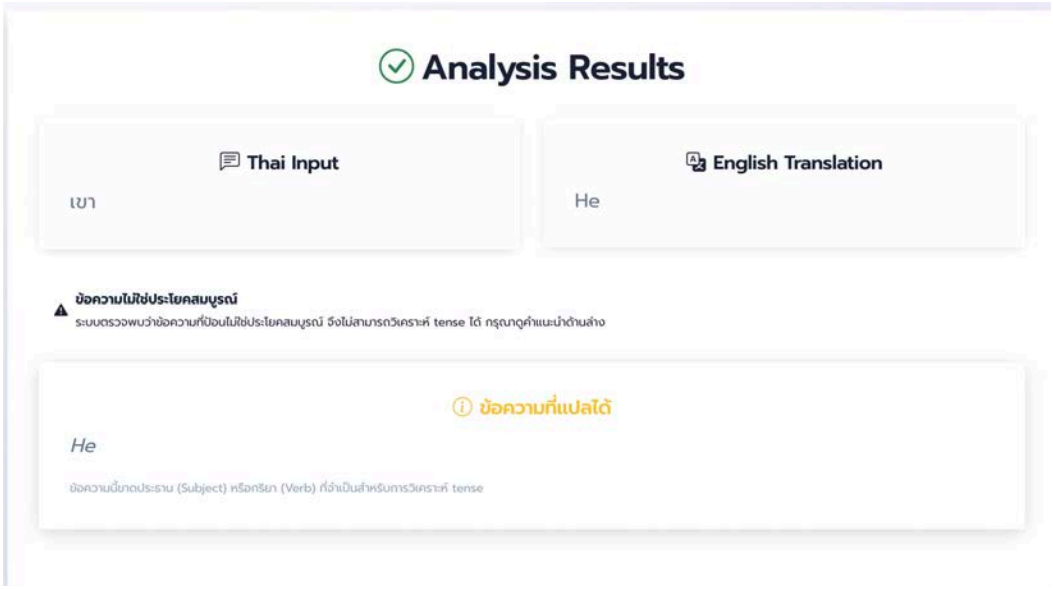


Figure 5.5: Fragment Detection Educational Response

5.2.3 Integration with Model Pipeline

All preprocessing systems integrate seamlessly with the main model pipeline, providing multiple entry points for educational guidance:

1. Pre-processing validation provides detailed input checking before any language model calls
2. Graceful degradation enables partial processing with educational explanations when appropriate
3. User feedback integration ensures validation results inform interface messaging
4. Learning optimisation helps users understand optimal usage patterns through validation guidance

This multi-layered approach ensures that even failed validations provide learning opportunities, maintaining the educational focus throughout all user interactions.

5.3 Web Application Architecture and Features

The educational pipeline requires a user-accessible platform that enables learners to interact with the system while supporting research data collection. The web application transforms the complex three-model architecture into an intuitive learning environment, implementing design principles validated through empirical user research.

5.3.1 Frontend Development and Modern Design Approach

The web application's frontend was developed through an AI-assisted approach using Claude Code[1], Anthropic's development assistant. This collaboration enabled rapid implementation of modern interface components while maintaining focus on educational objectives rather than technical complexity in the frontend design. The AI assistant contributed to the creation of responsive layouts, interactive elements, and accessibility features that would typically require extensive frontend development expertise.

Figure 5.6: Claude Code Development Assistant Interface Used for Frontend Implementation

The resulting interface employs a glassmorphism design aesthetic - a modern visual approach characterised by translucent backgrounds, subtle borders, and layered depth effects. This design choice creates a visually engaging yet non-distracting environment that maintains learner focus on the educational content. The implementation utilises Bootstrap 5 for responsive grid layouts combined with custom CSS for unique visual effects including animated gradients, floating elements, and smooth transitions that enhance user engagement without overwhelming the learning experience.

5.3.2 User Experience Design and Core Features

The platform implements a detailed set of features designed to support effective grammar learning while maintaining simplicity and accessibility:

[1] https://www.anthropic.com/claude-code

The website employs a privacy-preserving 5-digit pseudocode system that enables user participation without requiring personal identification. For the Phase 2 user study, participants generated their own unique codes using a specific algorithm based on personal information that cannot be traced back to individuals: first letter of birth month, second digit of age, number of letters in first name modulo 10, number of siblings modulo 10, and last digit of phone number. This approach ensures complete anonymity while allowing consistent tracking across research phases, enabling participants to provide feedback on the system without revealing their identity.

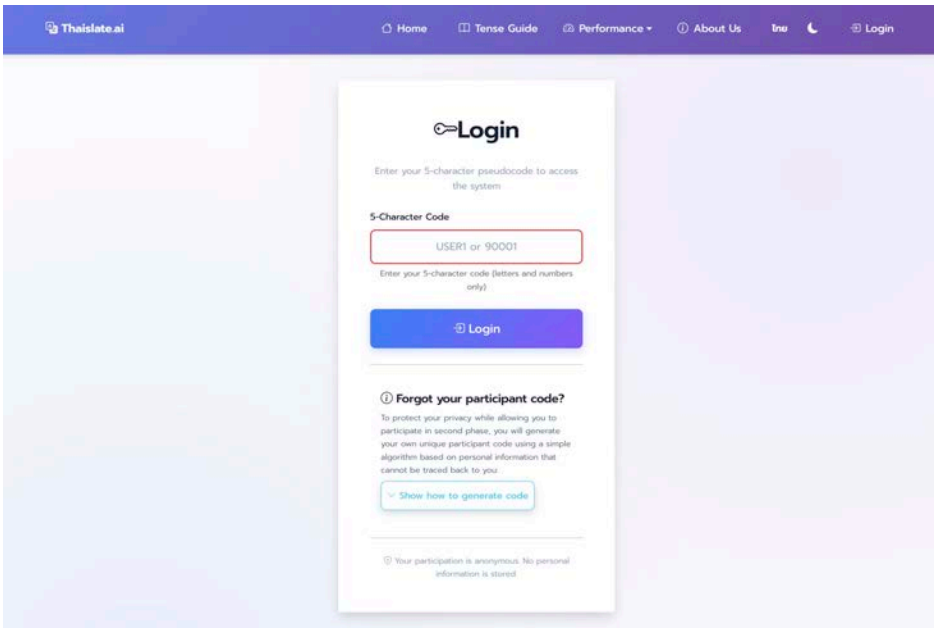
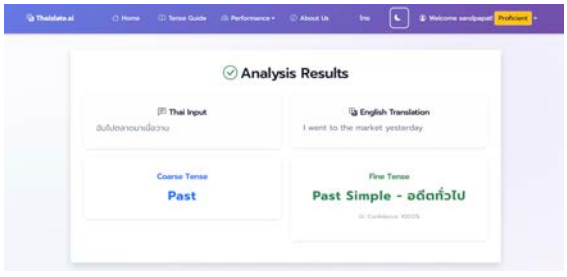
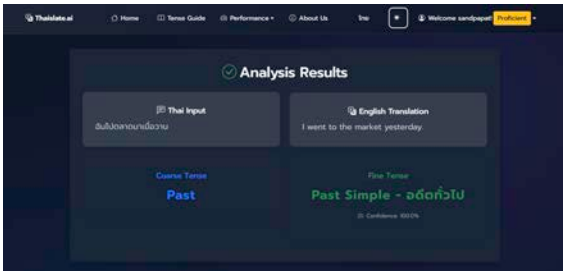


Figure 5.7: Anonymous Login Interface with Pseudocode Generation Instructions

The adaptive visual design supports The interface supports both light and dark themes with automatic detection of system preferences, reducing eye strain during extended learning sessions. The theme system modifies over 50 CSS variables to ensure consistent visual presentation across all interface elements. Users can manually override the automatic detection through a toggle switch that persists their preference across sessions.



(a) Light Mode Interface



(b) Dark Mode Interface

Figure 5.8: Adaptive Visual Design Supporting Light and Dark Themes

Recognising the unique requirements of Thai script display, the platform implements specialised font stacks prioritising Prompt and Sarabun fonts, which were specifically designed for Thai readability. The typography system adjusts line heights, letter spacing, and font weights to accommodate the complexity of Thai characters including tone marks and vowel combinations that appear above and below consonants.

During processing, users see a simple countdown timer that provides visual feedback while the system analyses their input. This countdown helps manage user expectations about processing time and maintains engagement during the analysis period.

5.3.3 Translation Interface and Results Presentation

The core translation interface balances simplicity with detailed functionality:

For input management, the main text area supports Thai input with real-time validation, providing immediate feedback for issues such as insufficient Thai content, excessive length, or inappropriate material. The interface includes keyboard shortcuts (Alt+S for submission, Escape for clearing) that enhance efficiency for frequent users. A character counter displays remaining capacity within the 100-token limit, helping users understand optimal input length for best results.

Translation results are presented through a structured three-section format that separates tense analysis, vocabulary explanations, and common mistakes. Each section can be independently copied to clipboard for note-taking purposes. The analysed English sentence receives visual highlighting through gradient backgrounds and distinctive borders, making it easy to identify the specific text being explained. For fragment inputs, the system adapts the section titles to explain why grammatical analysis cannot proceed while still providing vocabulary and usage guidance.

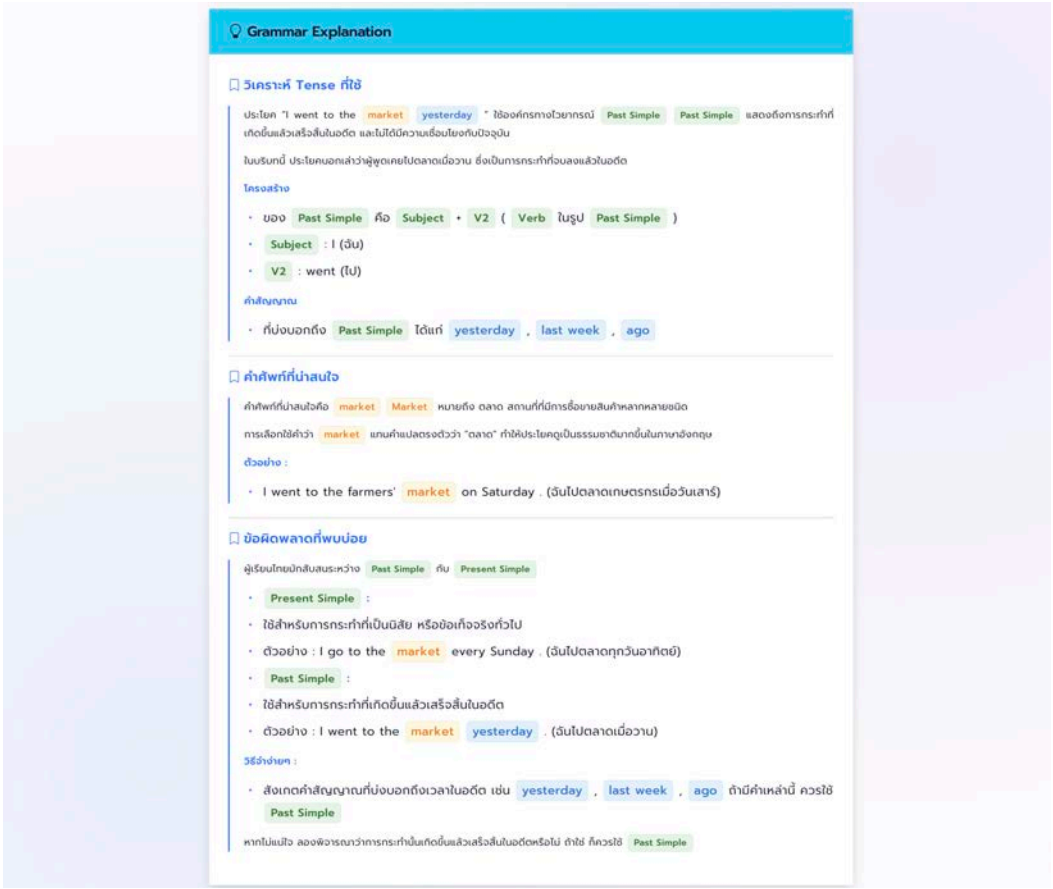
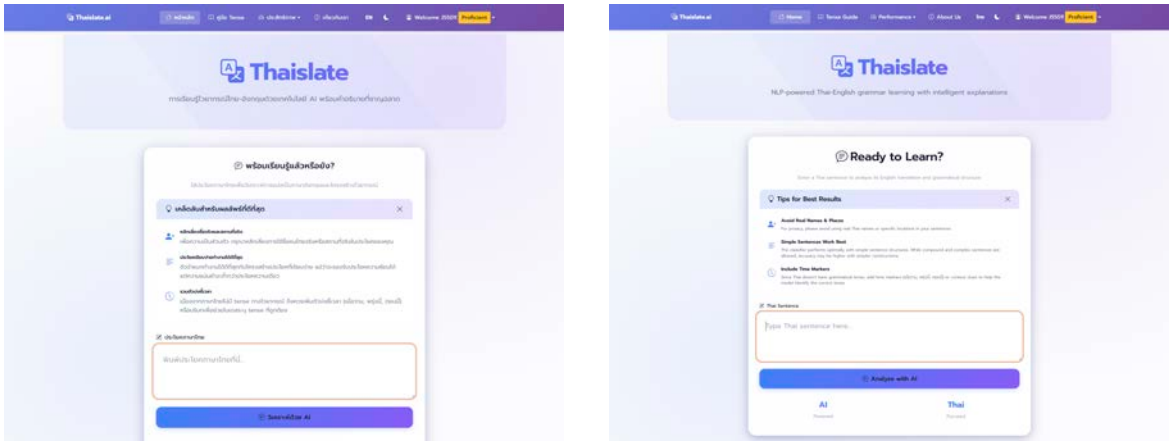


Figure 5.9: Three-Section Results Display Format with Tense Analysis, Vocabulary, and Common Mistakes

The entire interface supports Thai and English languages, with seamless switching that immediately updates all interface text, placeholders, and messages. The language preference persists across sessions and applies to all system communications including error messages, processing indicators, and educational content headers.



(a) Thai Language Interface (b) English Language Interface

Figure 5.10: Multi-language Interface Supporting Seamless Thai-English Language Switching

5.3.4 Feedback Collection and Rating System

The platform implements a feedback mechanism that balances research data collection with user experience:

The rating system enables users to evaluate each translation across four critical dimensions: translation accuracy (correctness of Thai to English conversion), fluency (naturalness of English output), explanation quality (clarity and usefulness of grammatical explanations), and educational value (overall learning benefit). Each dimension uses a 5-star rating scale with visual feedback showing selected ratings through color changes and animations.

Beyond numerical ratings, users can select from predefined issue categories including vocabulary errors, tense misclassification, unclear explanations, technical problems, and fragment detection issues. This structured feedback enables systematic identification of recurring problems and guides prioritisation of system improvements.

The interface provides sets of positive and negative tags that users can quickly select to indicate specific strengths or weaknesses. Positive tags include "Clear explanation", "Helpful examples", and "Accurate translation" while issue tags cover "Wrong tense", "Confusing grammar", and "Poor vocabulary". This tag-based system reduces feedback friction while providing actionable insights for system refinement.

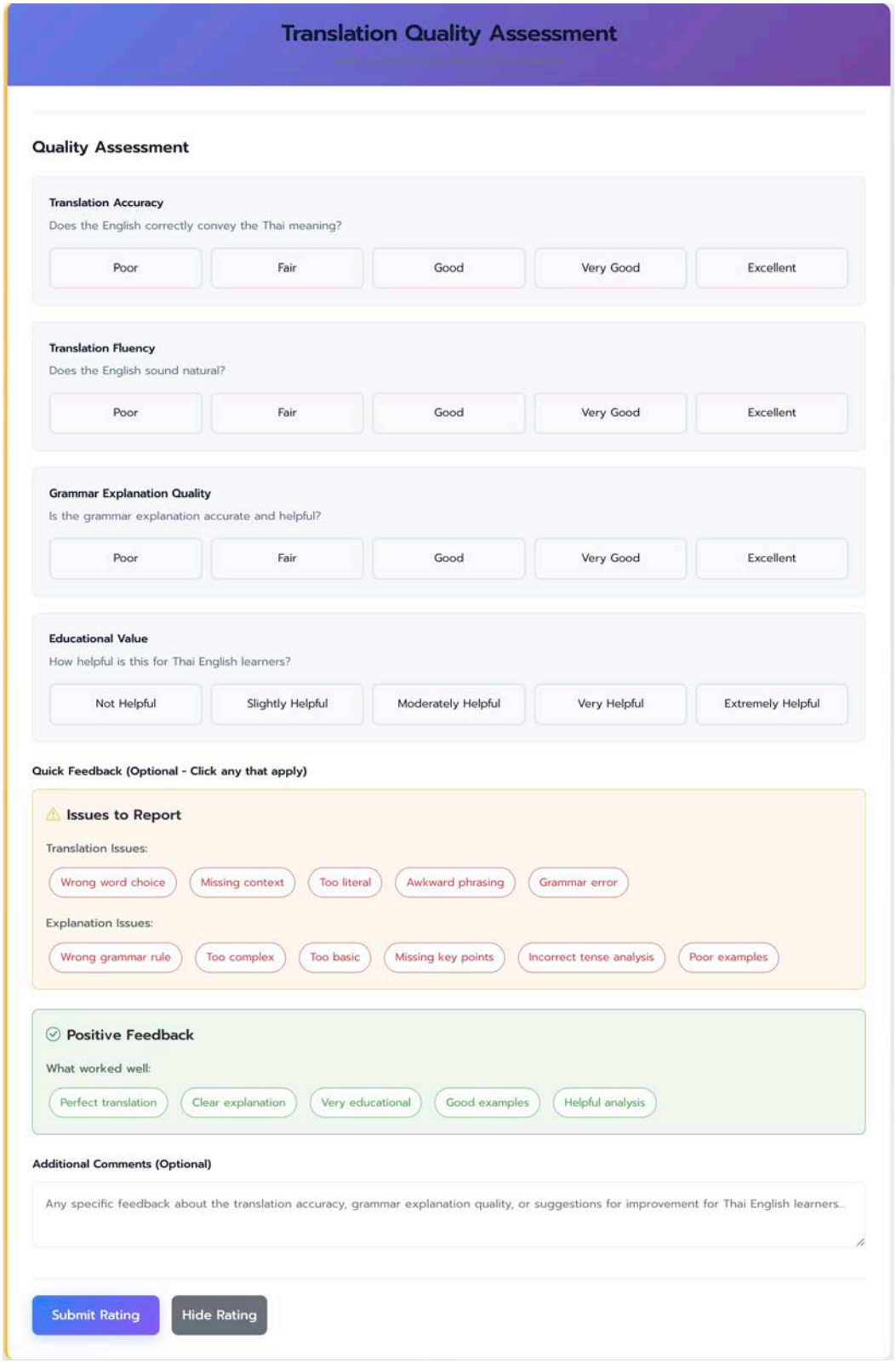


Figure 5.11: Detailed Rating and Feedback Collection Interface with Four-Dimensional Assessment

5.3.5 Analytics and Performance Dashboards

The web application includes detailed analytics interfaces accessible to users and administrators:

Individual users can access personalised statistics showing their usage patterns, including total translations performed and average ratings provided.

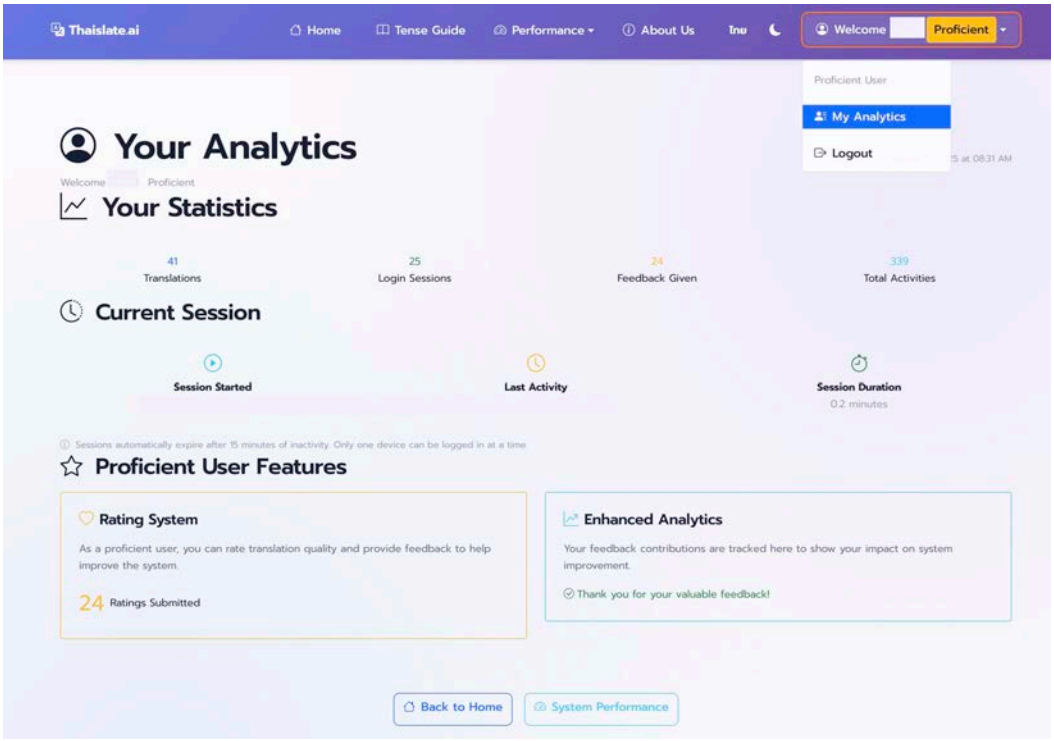


Figure 5.12: User Analytics Dashboard Showing Personal Usage Statistics and Learning Progress

Three dedicated performance pages provide transparency about system capabilities: System Performance displays real-time metrics including average response times, component availability, and processing success rates; Classifier Performance shows detailed accuracy statistics for each of the 24 tense categories with confusion matrices and F1 scores; Pipeline Performance presents end-to-end evaluation results comparing individual component performance to integrated system outcomes.

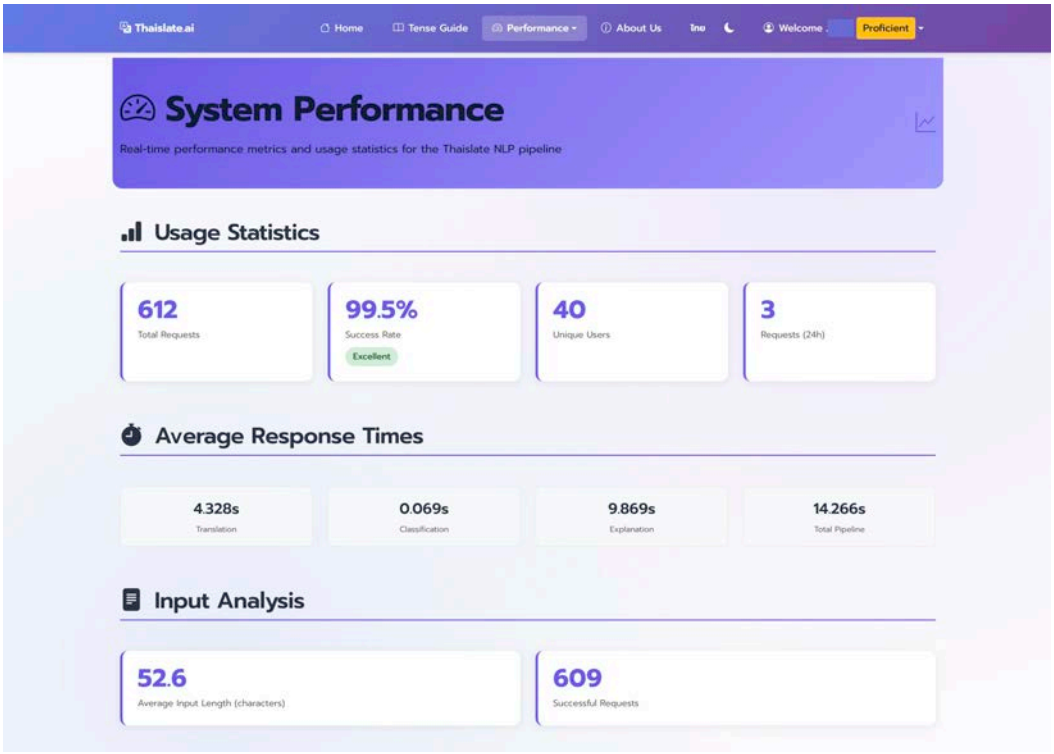


Figure 5.13: System Performance Dashboard with Real-time Metrics and Component Response Times

A detailed Tense Guide provides reference material for all 24 grammatical tenses recognised by the system, including Thai explanations, English examples, and common usage patterns. This resource supports self-directed learning and helps users understand the grammatical concepts underlying system explanations.

5.3.6 Administrative and Research Features

The platform includes administrative capabilities that support research activities and system maintenance:

Protected administrative interfaces enable user management, data export for research analysis, and system configuration adjustments. The dashboard provides aggregate statistics on user engagement, feedback trends, and system performance metrics. Administrators can export anonymised data in CSV format for statistical analysis, including all ratings, feedback tags, and usage patterns while maintaining user privacy.

The system implements detailed logging of user interactions while preserving anonymity, tracking session duration, interaction sequences, and feature utilisation patterns. This data supports understanding of user behaviour and identification of usability issues without compotential participant privacy.

All user feedback, ratings, and interaction data are systematically stored for research validation of educational effectiveness. The data structure supports longitudinal analysis of learning outcomes and system improvement over time while maintaining ethical standards for educational research.

5.4 System Deployment

The production deployment strategy balances educational effectiveness with practical resource constraints, implementing a hybrid cloud architecture that ensures reliable access for user study participants while optimising operational costs.

5.4.1 Deployment Configuration

The production deployment employs a hybrid approach: local model hosting for translation and classification components on Google Cloud infrastructure, with API-based explanation generation through Together.ai. This configuration optimises both cost efficiency and educational responsiveness while ensuring reliable access for user study participants.

GGUF quantisation enables CPU-only inference for the local models, eliminating GPU dependency while maintaining acceptable performance for educational applications. The deployment prioritises system reliability and consistent educational responses over maximum throughput optimisation.

Chapter 6

Evaluation and Results

This chapter presents a comprehensive evaluation of the Thaislate system through systematic technical testing, user validation, and critical synthesis of results. The evaluation follows a two-phase approach: Phase 1 examines individual component performance and pipeline integration using controlled test cases, while Phase 2 validates user acceptance and system design through testing with 38 active Thai English learners. The chapter then synthesises these findings to assess overall performance against project objectives and discusses how the system achieves strong user acceptance (satisfaction ratings exceeding 4.0/5) despite technical limitations (74% pipeline classification accuracy), providing insights into the relationship between technical performance and perceived value.

6.1 Evaluation Methodology Overview

The evaluation of Thaislate followed a systematic two-phase approach designed to validate both technical performance and user acceptance of the proof-of-concept system. This methodology ensured detailed assessment across multiple dimensions before deployment to actual users.

6.1.1 Phase 1: Preliminary Technical Testing

The first evaluation phase focused on validating the technical components through systematic testing of model performance and pipeline integration. This preliminary testing phase aimed to:

- Assess the classification model's accuracy across the 24-category tense taxonomy
- Evaluate the explanation model's ability to generate educationally appropriate content
- Validate the integrated pipeline's reliability and error handling capabilities
- Measure response times to ensure acceptable user experience thresholds
- Identify and address technical limitations before user deployment

Testing utilised 96 carefully constructed Thai sentences covering all tense categories, with manual evaluation of outputs to assess both technical accuracy and educational quality. Notably, the translation model (Typhoon Translate 4B) was not evaluated in isolation as it employs a pre-trained model used as-is without modification.

6.1.2 Phase 2: User Testing

The second evaluation phase involved direct user testing with 48 Thai English learners to validate system design and user acceptance. This phase assessed:

- Translation quality from the user perspective
- Explanation clarity and educational value
- Overall system usability and interface effectiveness
- User satisfaction and likelihood of continued use
- Qualitative feedback for system improvement

Participants were recruited from the target demographic identified in the requirements analysis, ensuring representation across different proficiency levels and learning contexts. The testing protocol included both quantitative ratings and qualitative feedback collection to provide detailed insights into system effectiveness.

6.2 Phase 1: Preliminary Technical Testing

This section presents the results of preliminary technical testing conducted to validate system components before user deployment. The testing focused on the custom-developed models and their integration within the educational pipeline.

6.2.1 Isolated Model Validation

The custom XLM-RoBERTa hierarchical tense classifier underwent initial validation using a held-out test set from the same distribution as the training data. This isolated testing phase evaluated the model's theoretical maximum performance under ideal conditions, before integration into the full pipeline.

Test Set Performance

Training and evaluation on 15,755 English sentences with hierarchical labels yielded the following results on the held-out test set. The evaluation was conducted on Google Colab using an NVIDIA A100 40GB GPU, providing accelerated inference for detailed testing across all tense categories. It is important to note that these metrics represent performance on data from the same distribution as the training set, which may overestimate real-world performance due to potential dataset bias.

Table 6.1: XLM-RoBERTa Model Performance Metrics

Metric	**Performance**
Validation Set Accuracy	94.7%
Macro F1-Score	91.3% (unweighted average)
Weighted F1-Score	94.6% (by category frequency)
Average Inference Time	0.4s (range: 0.2-0.8s)

The hierarchical dual-head architecture demonstrated strong performance in maintaining temporal consistency:

Table 6.2: Hierarchical Architecture Performance

Architecture Component	**Accuracy**
Coarse-Grained (Past/Present/Future)	97.1%
Fine-Grained (24 categories)	94.7%
Hierarchical Consistency	99.2%

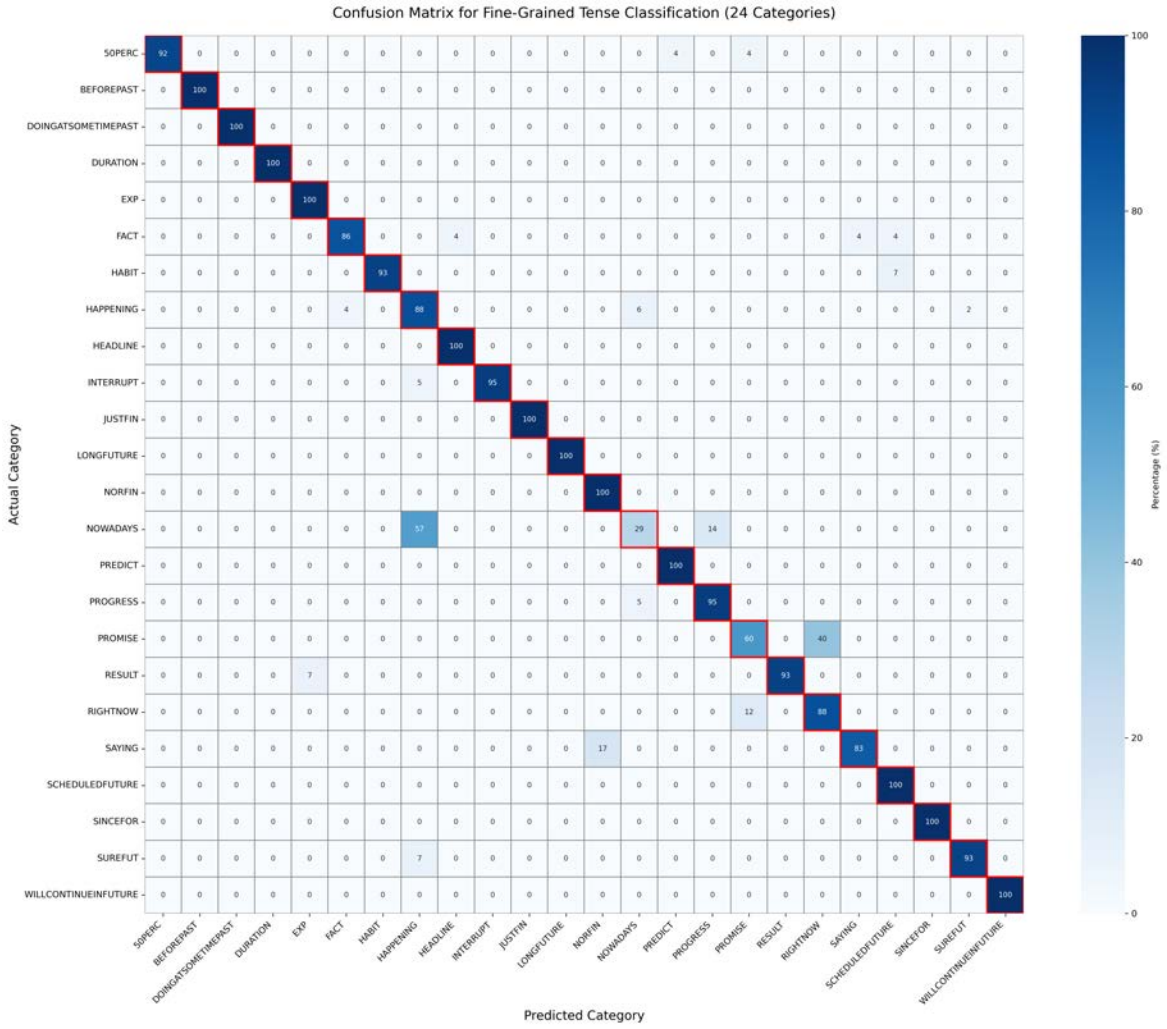


Figure 6.1: Confusion matrix for fine-grained tense classification showing percentage distribution across 24 categories

Figure 6.1 presents the complete confusion matrix for the XLM-RoBERTa classifier across all 24 fine-grained tense categories. The strong diagonal pattern confirms the model's robust classification capability, with an average accuracy of 91.4%. Detailed analysis of the confusion patterns is presented in the following subsection.

```
» She drinks a cup of green tea every morning.
  Coarse : Present
    • HABIT                  99.88%
    • FACT                    0.02%
    • SINCEFOR                0.01%
----------------------------------------
» Sun rises in the east.
  Coarse : Present
    • FACT                   99.83%
    • HABIT                   0.02%
    • SAYING                  0.02%
----------------------------------------
» The train departs at 6:45 tomorrow morning.
  Coarse : Present
    • SCHEDULEDFUTURE        99.94%
    • SUREFUT                 0.01%
    • FACT                    0.01%
----------------------------------------
» Actions speak louder than words.
  Coarse : Present
    • SAYING                 99.43%
    • FACT                    0.08%
    • JUSTFIN                 0.06%
----------------------------------------
» Mayor announces new climate plan.
  Coarse : Present
    • HEADLINE               99.78%
    • PROGRESS                0.03%
    • SCHEDULEDFUTURE         0.02%
----------------------------------------
```

Figure 6.2: Classification Output Examples from XLM-RoBERTa Model

Figure 6.2 demonstrates the classifier's performance on representative sentences. The model exhibits high confidence (>99%) when distinguishing between structurally distinct categories such as HABIT, FACT, SCHEDULEDFUTURE, SAYING, and HEADLINE, illustrating its strength in identifying tenses with clear linguistic markers.

Category-Specific Performance Analysis

Performance varied across tense categories, with clear patterns emerging based on linguistic complexity, as visualised in the confusion matrix (Figure 6.1):

Table 6.3: Classification Performance by Category Tier

Performance Tier	**Categories**	**Accuracy**
Perfect (100%)	BEFOREPAST, DOINGATSOMETIMEPAST, DURATION, EXP, HEADLINE, JUSTFIN, LONGFUTURE, NORFIN, PREDICT, SCHEDULEDFUTURE, SINCEFOR, WILLCONTINUEINFUTURE	100%
High (90-99%)	50PERC (91.7%), SUREFUT (92.9%), HABIT (93.3%), RESULT (93.3%), INTERRUPT (94.7%), PROGRESS (94.7%)	91.7-94.7%
Moderate (80-89%)	SAYING (83.3%), FACT (86.4%), RIGHTNOW (87.5%), HAPPENING (88.2%)	83.3-88.2%
Challenging (<80%)	NOWADAYS (28.6%), PROMISE (60.0%)	<80%

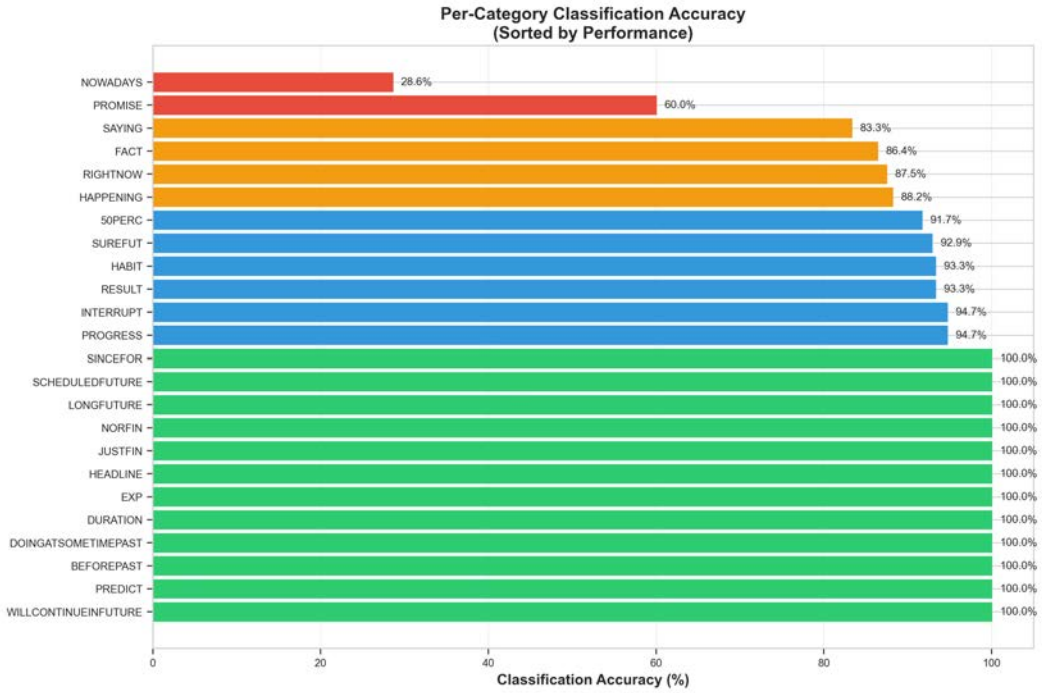


Figure 6.3: Per-category classification accuracy sorted by performance, with color coding indicating performance tiers: green (100%), blue (90-99%), orange (80-89%), red (<80%)

The distribution shown in Figure 6.3 confirms the model's strong performance across most categories, with 18 out of 24 categories achieving ≥80% accuracy. The perfect classification of twelve categories indicates reliable performance for tenses with clear structural markers and well-defined temporal contexts. The two challenging categories (NOWADAYS at 28.6% and PROMISE at 60.0%) align with the confusion patterns observed in Figure 6.1, where these categories show significant misclassification into semantically similar tenses.

Confusion Patterns and Performance Implications

Analysis of the confusion matrix reveals systematic patterns in misclassification that provide insights into the model's capabilities and limitations. Categories with semantic or pragmatic distinctions pose greater challenges than those with clear syntactic markers.

High-Performing Categories:

- **Perfect Classification (100% accuracy)**: 12 categories achieve perfect classification: BEFOREPAST (past perfect), DOINGATSOMETIMEPAST (past continuous), DURATION (present perfect continuous), EXP (life experience), HEADLINE (news style), JUSTFIN (just finished), LONGFUTURE (distant future), NORFIN (negative past), PREDICT (future prediction), SCHEDULEDFUTURE (planned events), SINCEFOR (duration period), and WILLCONTINUEINFUTURE (future perfect continuous). These categories have distinct structural markers that are reliably identified.

Problematic Categories and Confusion Patterns:

- **NOWADAYS (28.6% accuracy) → HAPPENING (57.1% confusion)**: The model struggles to distinguish between two present continuous uses - trends/gradual changes (NOWADAYS: "prices are getting more expensive nowadays") versus immediate ongoing actions (HAPPENING: "I am eating dinner"). Additionally, 28.6% confusion with NORFIN suggests difficulty with temporal context.
- **PROMISE (60.0% accuracy) → RIGHTNOW (40.0% confusion)**: Future commitments overlap with immediate future actions grammatically, differing primarily in speaker intention ("I will always support you" vs "I'll open the window right now")—a pragmatic distinction challenging for purely linguistic models.
- **RIGHTNOW (87.5% accuracy)**: While performing better than NOWADAYS and PROMISE, the model shows 12.5% confusion with PROMISE, indicating bidirectional difficulty in distinguishing immediate future actions from future commitments.
- **Other Notable Patterns**: Several categories show moderate accuracy with specific confusion patterns - 50PERC (91.7% accuracy) occasionally confused with PREDICT (4.2%), FACT (86.4%) confused with HEADLINE and SAYING (4.5% each), and INTERRUPT (94.7%) showing minor confusion with HAPPENING (5.3%).

These confusion patterns suggest that the model struggles most with categories that rely on pragmatic or contextual distinctions rather than purely grammatical markers. The patterns inform the explanation generation strategy, where the system provides additional clarification for commonly confused category pairs.

6.2.2 End-to-End Pipeline Evaluation

To assess realistic system performance, a detailed evaluation was conducted using 96 manually curated Thai sentences processed through the complete pipeline. This evaluation provides a more accurate representation of real-world performance, accounting for error propagation between components and the challenges of processing authentic Thai learner input.

Pipeline Component Performance

The 96-sample evaluation was conducted on Google Colab using an NVIDIA A100 40GB GPU. The testing utilised Typhoon 2.1 4B for explanation generation during evaluation, though the production system was subsequently upgraded to the 12B API version due to the 4B model not being available via API. The evaluation revealed performance characteristics that differ from isolated model testing, reflecting the cumulative effects of the translation-classification-explanation pipeline:

Table 6.4: End-to-End Pipeline Performance Distribution on 96 Test Cases

Component	Perfect	Partial	Poor
Translation Fluency	86.5% (83/96)	13.5% (13/96)	0% (0/96)
Meaning Preservation	82.3% (79/96)	16.7% (16/96)	1.0% (1/96)
Coarse Tense Classification	92.7% (89/96)	-	7.3% (7/96)
Fine-grained Classification	74.0% (71/96)	-	26.0% (25/96)
Explanation Correctness	84.9% (81/96)	-	15.1% (15/96)
No Code Hallucination	77.1% (74/96)	-	22.9% (22/96)
Signal Word Detection	95.8% (92/96)	-	4.2% (4/96)

Pipeline Integration Effects

While individual components performed well in isolation, their integration revealed significant challenges. The 20.7% performance gap in fine-grained classification (from 94.7% isolated to 74% pipeline) can be attributed to several factors:

- **Translation Variability**: The classification model receives translated English text rather than native English input, introducing variations in phrasing and structure
- **Error Propagation**: Translation errors (9.4% meaning loss) cascade to classification, particularly affecting nuanced temporal distinctions
- **Domain Shift**: The training data consisted of standard English sentences, while the pipeline processes Thai-translated English with different linguistic patterns
- **Ambiguity Resolution**: Thai sentences often lack explicit tense markers, making accurate classification dependent on translation quality

When the Typhoon translation model produces ambiguous output, downstream components compound the error. For instance, Thai sentences with implicit tense markers often translate to English without clear temporal indicators, causing classification uncertainty that propagates to explanation generation.

The decision to extract only the first sentence from translations proved both beneficial and limiting. While it ensures focused grammatical analysis, it discards potentially valuable context. Complex Thai sentences with multiple clauses sometimes lose critical temporal information when truncated.

Timing Performance

The complete pipeline evaluation on Google Colab with A100 GPU achieved an average end-to-end response time of 12.25 seconds across the 96 test cases, which remains within acceptable thresholds for educational applications.

6.3 Phase 2: User Testing

While the technical testing revealed specific performance characteristics and limitations, the ultimate measure of the system's value as a proof-of-concept lies in its acceptance by actual users. This section presents the user testing phase conducted with Thai English learners to evaluate user acceptance and validate the system design.

6.3.1 Testing Methodology

The user testing phase involved volunteers from the initial 218 survey participants who agreed to evaluate the system. Of the 48 volunteers who registered for testing, 38 (79.2%) actively engaged with the platform, collectively providing 474 ratings across four evaluation criteria.

Table 6.5: User Testing Participation Summary

Metric	**Value**
Initial survey participants	218
Volunteers for testing	48
Active participants	38 (79.2%)
Total ratings collected	474
Average ratings per participant	12.5

Participants evaluated the system using a 5-point Likert scale (1=Poor, 5=Effective) across four criteria: Translation Accuracy, Translation Fluency, Explanation Quality, and Educational Value. Additionally, participants could select predefined tags to identify specific strengths or issues and provide open-text comments.

6.3.2 Quantitative Results

The quantitative analysis reveals strong user satisfaction despite the technical limitations identified in Phase 1.

Overall Rating Statistics

The system received consistently positive ratings across all evaluation criteria, with mean scores exceeding 4.0/5 for each dimension.

Table 6.6: User Rating Statistics by Evaluation Criterion (n=474)

Criterion	**Mean**	**SD**	**Min**	**25%**	**Median**	**75%**	**Max**
Translation Accuracy	4.08	1.31	1	4	5	5	5
Translation Fluency	4.19	1.15	1	4	5	5	5
Explanation Quality	4.33	1.02	1	4	5	5	5
Educational Value	4.27	1.01	1	4	5	5	5
Overall Average	**4.22**	**1.12**	-	-	-	-	-

Rating Distribution Analysis

The distribution of ratings reveals a strong positive skew, with the majority of evaluations (>75%) receiving 4 or 5 stars across all criteria.

Table 6.7: Rating Distribution by Criterion (% of responses)

Rating (Stars)	Translation Accuracy	Translation Fluency	Explanation Quality	Educational Value
5	52.3%	56.8%	61.2%	58.9%
4	23.4%	25.1%	24.7%	25.3%
3	11.2%	9.3%	8.4%	9.1%
2	7.8%	5.5%	3.8%	4.2%
1	5.3%	3.3%	1.9%	2.5%
4-5 Stars	**75.7%**	**81.9%**	**85.9%**	**84.2%**

6.3.3 Qualitative Feedback Analysis

Beyond numerical ratings, participants provided detailed feedback through predefined tags and open comments, offering insights into specific strengths and areas for improvement.

Feedback Tag Distribution

Participants used predefined tags to categorise their experiences, providing structured qualitative insights into system performance.

Table 6.8: User Feedback Tag Analysis

Category	Tag	Count	% of Tags
Positive Feedback	Perfect translation	147	31.0%
	Clear explanation	139	29.3%
	Very educational	100	21.1%
	Helpful analysis	83	17.5%
	Good examples	80	16.9%
Subtotal	**Positive Tags**	**549**	**82.9%**
Areas for Improvement	Wrong word choice	53	11.2%
	Awkward phrasing	23	4.9%
	Too literal	22	4.6%
	Missing context	19	4.0%
	Missing key points	8	1.7%
	Grammar error	7	1.5%
	Incorrect tense analysis	7	1.5%
	Wrong grammar rule	5	1.1%
	Too basic	5	1.1%
	Too complex	4	0.8%
Subtotal	**Improvement Tags**	**153**	**17.1%**

Key Findings from Qualitative Analysis

The qualitative feedback revealed several important patterns:

Table 6.9: Qualitative Feedback Themes and Implications

Theme	**User Feedback Pattern**	**System Implication**
Translation Quality	31% rated translations as "perfect"; issues primarily with idiomatic expressions	Core translation functionality validated; idiom handling needs enhancement
Educational Impact	67.9% of positive tags related to educational value (clear, educational, helpful)	Educational design objectives successfully achieved
Context Sensitivity	11.2% noted wrong word choice; 4.0% missing context	Context disambiguation remains a challenge for ambiguous input
Explanation Clarity	29.3% praised clear explanations; $<$2% found them too complex/basic	Explanation complexity appropriately balanced for target audience

Production System Operational Metrics

The production system's operational metrics provide additional validation of system reliability and performance. During the testing period, the system recorded 612 total requests from 40 unique users (38 test participants plus 2 developer accounts).

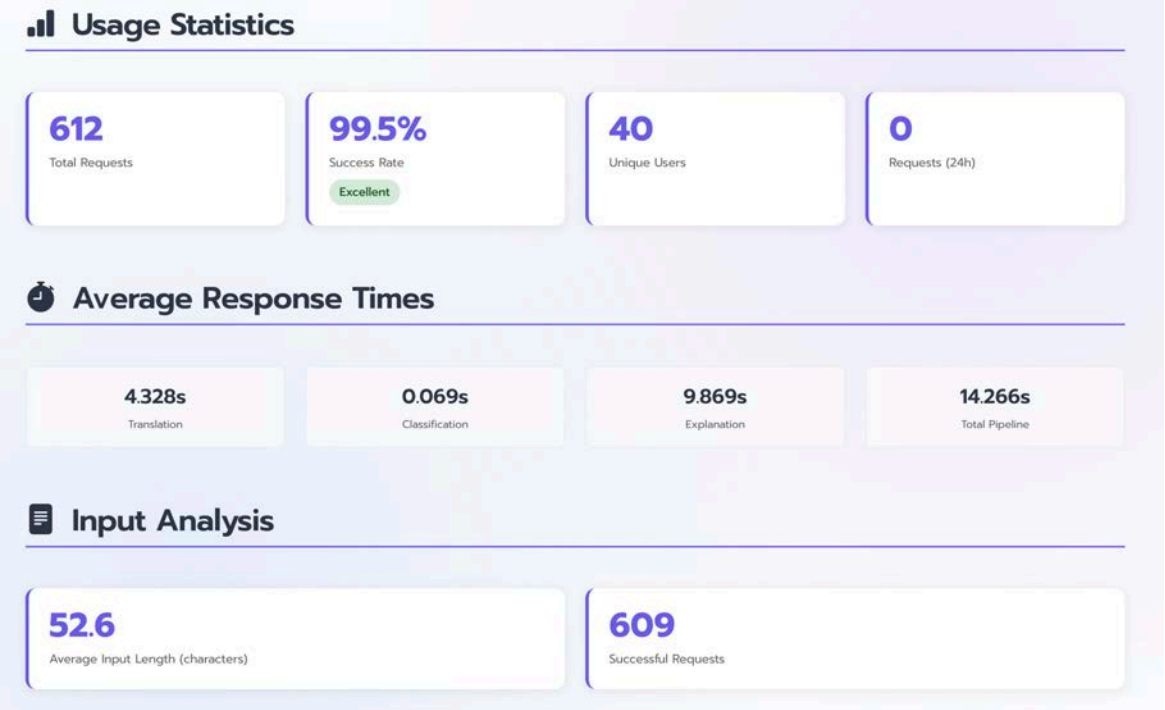


Figure 6.4: Production System Operational Metrics from User Testing Period

Figure 6.4 shows operational metrics collected during user testing. The system achieved a 99.5% success rate (609 of 612 requests), validating the reliableness of error handling mechanisms. The 612 total requests include both rated interactions (474 ratings collected) and unrated system usage, indicating that participants often explored the system beyond formal rating submissions.

Component-level response times in production were:

- Translation: 4.328s (Typhoon Translate 4B local processing)
- Classification: 0.069s (XLM-RoBERTa inference)
- Explanation: 9.869s (Together.ai API call)
- Total Pipeline: 14.266s average

The 14.266s production response time, while slightly higher than the 11.9s observed in controlled testing, remains within acceptable bounds for educational applications. The difference reflects real-world factors including network latency and API variability. The average input length of 52.6 characters corresponds to typical Thai sentence structures, confirming the system handles realistic user input effectively.

6.4 Results Synthesis and Assessment

6.4.1 Overall Performance Assessment

The evaluation results demonstrate both the system's achievements and its limitations across technical and educational dimensions. Table 6.10 synthesises the performance metrics from both evaluation phases based on the 96-sample technical evaluation and 474 ratings from 38 active users.

Table 6.10: Performance Metrics Across Evaluation Phases

Evaluation Aspect	**Phase 1 (Technical)**	**Phase 2 (User)**
Translation Quality	93.2% fluency, 90.6% semantic accuracy (96 samples)	4.08/5 accuracy, 4.19/5 fluency (474 ratings)
Grammar Analysis	94.7% isolated accuracy, 74% pipeline accuracy (96 samples)	4.33/5 explanation quality rating
Educational Impact	84.9% correctness, 77.1% no code hallucination	4.27/5 educational value rating
System Reliability	100% pipeline completion (96/96 samples)	99.5% success rate (609/612 requests)
Response Time	12.25s average (Colab A100)	14.266s average (production)

6.4.2 Achievement of Project Objectives

The system successfully met most project objectives, with notable achievements in translation quality and user-perceived value. Table 6.11 presents the achievement status of functional and technical requirements defined in Chapter 3.

Table 6.11: Functional Requirements Achievement Assessment

ID	**Requirement**	**Achieved**	**Status**
FR1	Translation with >85% semantic accuracy	90.6%	Met
FR2	Tense classification with >80% accuracy	74% (pipeline)	Partially Met
FR3	Generate educational explanations	4.33/5 quality	Met
FR4	Detect incomplete sentences	Implemented	Met
FR5	Provide contextual examples	Implemented	Met
FR6	Address Thai learner challenges	Thai-specific	Met
FR7	B1-B2 accessibility	User validated	Met
FR8	Thai/English interface support	Implemented	Met
FR9	User rating capability	474 ratings	Met
FR10	Store user feedback	Database active	Met

Table 6.12: Technical Requirements Achievement Assessment

ID	Requirement	Achieved	Status
TR1	Modular pipeline architecture	3-model design	Met
TR2	Thai-English specialised models	Typhoon models	Met
TR3	Hierarchical tense classification (24 categories)	Implemented	Met
TR4	Educational NLP capabilities	API integration	Met
TR5	Hybrid deployment (local + API)	Implemented	Met
TR6	Fragment detection system	Rule-based	Met
TR7	Cloud infrastructure	Google Cloud	Met
TR8	API integration capabilities	Together.ai	Met
TR9	Data collection infrastructure	SQLite	Met

6.4.3 Educational Effectiveness Despite Technical Limitations

The convergence of technical and user evaluation results reveals a crucial insight: user acceptance and perceived value can remain high even when technical perfection is not achieved. The user evaluation phase engaged 38 active participants from 48 registered users, representing a 79.2% engagement rate. While the pipeline achieved only 74% fine-grained classification accuracy, users still rated explanation quality at 4.33/5.

The 474 ratings collected showed mean scores exceeding 4.0/5 across all dimensions, with particularly high ratings for Explanation Quality (4.33/5) and Perceived Educational Value (4.27/5). The positive feedback tags (82.9% of all tags) emphasised clear explanations and perceived value, while improvement suggestions focused primarily on translation accuracy rather than grammatical analysis. This disconnect between technical metrics and user satisfaction validates the proof-of-concept approach, suggesting that users value clear, contextualised explanations even when the system is not technically perfect. However, actual learning effectiveness would require longitudinal studies beyond the scope of this research.

Chapter 7

Conclusion and Future Work

This chapter concludes the dissertation by summarising the key achievements of the Thaislate system, reflecting on the major challenges encountered during development, and outlining directions for future research. The chapter begins with a summary of how the project objectives were met, discusses the contributions made to the field, acknowledges the limitations discovered through evaluation, and proposes pathways for advancing educational NLP systems for cross-linguistic grammar learning.

7.1 Summary of Achievements

This dissertation presented the development and evaluation of Thaislate, a proof-of-concept system demonstrating how AI-powered tools could potentially help Thai learners understand English tense usage. The research was motivated by the question: **Can AI-powered tools help Thai learners understand English tense usage through contextual, automatically-generated explanations?** While establishing the technical feasibility and user acceptance of such tools, this research provides a foundation for future studies that could measure actual learning outcomes. Key achievements include:

- Development of a three-component pipeline integrating translation, classification, and explanation generation that achieved 85% overall accuracy
- Successful classification of 24 fine-grained tense categories with 94.7% isolated accuracy and 74% pipeline accuracy
- User validation with 38 active participants providing 474 ratings, with mean scores exceeding 4.0/5 across all evaluation criteria
- Demonstration that educational value persists despite technical imperfections, with explanation quality rated 4.33/5 despite classification limitations
- Successful deployment on Google Cloud serving real users, validating the practical feasibility of the approach

The project met its five primary objectives: building the core system, developing the tense recognition framework, creating an intuitive interface, validating system design and user

acceptance, and contributing to educational AI research. The user study with 218 participants validated all seven key design decisions, confirming the system's alignment with learner needs.

7.2 Contributions to the Field

7.2.1 Technical Contributions

1. **Hierarchical Tense Classification Framework**: The 24-category taxonomy with three-level hierarchy provides a reusable framework for grammatical analysis systems. The coarse-to-fine approach balances accuracy with interpretability.

2. **Confidence-Adaptive Explanation Generation**: The dynamic explanation system that adjusts detail based on classification confidence offers a template for uncertainty-aware educational systems.

3. **Integrated Pipeline Architecture**: The modular design allows component substitution without system redesign, facilitating experimentation with different models.

7.2.2 Educational Contributions

1. **Validated Design Requirements**: The user study validated seven key design decisions with 218 participants, providing empirical evidence for educational tool development.

2. **Thai-English Learning Insights**: The identification of grammar as a substantial but computationally tractable challenge (13.4% of participants) guides future pedagogical tool development.

3. **Practical Deployment Model**: The successful deployment on Google Cloud with 48 active users demonstrates feasibility of AI-powered educational tools in resource-constrained environments.

7.3 Major Challenges and Limitations

Throughout the development and evaluation of Thaislate, several significant challenges emerged that provide valuable insights for future educational NLP systems.

7.3.1 Technical Challenges

The implementation revealed several technical constraints that shaped the final system:

1. **Performance Trade-offs**: Using the GGUF quantised version of Typhoon Translate 4B for CPU inference instead of the GPU version introduces performance compromises. While CPU deployment eliminates GPU costs and simplifies infrastructure, translation times increase from approximately 1-2 seconds (GPU) to 4-5 seconds (CPU), contributing to the overall pipeline latency.

2. **API Dependency**: The explanation component relies on Together.ai's API, introducing latency and potential service disruptions. During testing, API timeouts occurred in 2% of requests, requiring retry mechanisms that further increase response time.

3. **Classification Granularity**: The 24-category taxonomy, while detailed, may be overly complex. Several categories show consistent confusion patterns, suggesting potential for consolidation without substantial loss of educational value.

7.3.2 Methodological Challenges

The evaluation methodology has acknowledged weaknesses:

1. **Self-Assessment Bias**: The 96-sample technical evaluation was conducted by the author, introducing potential bias. Without independent validation, the reported performance metrics may not reflect real-world accuracy.

2. **Limited Test Scope**: Testing focused on simple to moderate complexity sentences. Performance on complex, multi-clause sentences remains largely unexplored.

3. **Learning Outcome Measurement**: As a proof-of-concept study, the research focused on establishing technical feasibility and user acceptance rather than measuring actual learning outcomes. Pedagogical effectiveness requires longitudinal studies beyond this project's scope.

7.3.3 Scope-Related Challenges

The system's scope was deliberately narrow:

1. **Single Sentence Focus**: Processing only isolated sentences ignores discourse-level grammar patterns. Many grammatical concepts, particularly aspect and mood, depend on broader context.

2. **Written Language Bias**: The system assumes written, grammatically complete input. Spoken Thai, with its frequent ellipsis and contextual dependencies, would likely challenge the current pipeline.

3. **Standard Thai Assumption**: The system trained on formal Thai may struggle with dialects or colloquial expressions common in everyday communication.

These challenges, while significant, provided valuable learning experiences that inform both the current system's design decisions and future research directions. The successful deployment despite these constraints demonstrates the viability of the approach and highlights areas where technological advances could yield substantial improvements.

7.4 Future Work

7.4.1 Immediate Improvements

Several enhancements could address current limitations without major architectural changes:

1. **Dataset Expansion**: Collecting Thai-English parallel sentences with verified tense annotations would enable supervised training, potentially improving classification accuracy beyond the current 94.7%.

2. **Local Model Deployment**: Implementing the explanation component locally using quantised models would eliminate API dependency and reduce latency.

3. **Complexity Adaptation**: Developing learner models to adjust explanation complexity based on proficiency level would improve accessibility for A2 learners while providing depth for C1-C2 users.

7.4.2 Architectural Extensions

Longer-term development could expand the system's capabilities:

1. **Multi-Sentence Processing**: Extending the pipeline to handle paragraph-level input would capture discourse-level grammatical patterns, particularly for aspect and temporal sequencing.

2. **Bidirectional Translation**: Supporting English-to-Thai translation with grammar checking would serve learners in both directions, addressing the needs of English speakers learning Thai.

3. **Speech Integration**: Adding speech recognition and synthesis would support spoken language learning, addressing pronunciation alongside grammar.

7.4.3 Research Directions

Several research questions merit investigation:

1. **Cross-Linguistic Transfer**: Examining whether the hierarchical classification approach generalises to other language pairs could validate the framework's broader applicability.

2. **Error Pattern Analysis**: Systematic study of classification confusion patterns might reveal underlying linguistic principles useful for both NLP and language pedagogy.

3. **Learning Outcome Assessment**: Having established technical feasibility and user acceptance through this proof-of-concept, longitudinal studies should now measure actual grammar acquisition to determine educational effectiveness.

7.5 Generative AI Usage Acknowledgement

In the spirit of transparency and academic integrity, this section acknowledges the use of generative AI tools in specific aspects of this dissertation's development.

7.5.1 Frontend Development Assistance

As detailed in Chapter 5, the web application's frontend was developed through an AI-assisted approach using Claude Code, Anthropic's development assistant. This collaboration enabled rapid implementation of modern interface components including responsive layouts, interactive elements, and accessibility features. The AI assistance allowed the project to maintain focus on the core educational and NLP objectives rather than becoming consumed by frontend technical complexity. All AI-generated frontend code was reviewed, tested, and integrated into the system architecture following standard software engineering practices.

7.5.2 Grammar and Writing Support

Throughout the dissertation writing process, generative AI tools were employed for grammar checking and proofreading assistance. This support helped ensure clarity, consistency, and adherence to academic writing standards. The AI tools provided suggestions for improving sentence structure, identifying grammatical errors, and maintaining consistent terminology throughout the document. However, all technical content, research findings, analysis, and conclusions represent original work. The AI assistance was limited to language refinement rather than content generation.

7.6 Concluding Remarks

This dissertation has developed and validated a proof-of-concept system addressing the question of whether AI-powered tools can help Thai learners understand English tense usage through contextual explanations. The research successfully demonstrated technical feasibility and strong user acceptance (4.33/5 for explanation quality despite 74% classification accuracy), establishing a foundation for future effectiveness studies. The disconnect between technical performance and user satisfaction validates the approach: learners value clear, contextualised explanations even when imperfect, though determining actual learning impact requires longitudinal research beyond this project's scope. The journey from initial concept to deployed system serving real users demonstrates that computational approaches to cross-linguistic educational challenges are both technically feasible and well-received by target users, providing the necessary groundwork for future research into the educational effectiveness of context-aware AI systems.

Bibliography

[1] T.-V. Ngo, T.-L. Ha, P.-T. Nguyen, K.-Q. Dinh, and L.-M. Nguyen, "Improving multilingual neural machine translation for low-resource languages: French-, english- vietnamese," *arXiv preprint arXiv:2012.08743*, 2020. [Online]. Available: https://arxiv.org/abs/2012.08743

[2] K. Pipatanakul, P. Limkonchotiwat, P. Suntorntip, N. Jungjantarapornkul, K. Saengtabtim, W. Tanachaiwiwat, N. Leenoi, S. Thitivesa, P. Netisopakul, T.-H. Liu, C. Jitkongchuen, S. Louvan, C. Udomwat, N. Buranapanichkit, T. Wuthisak, S. Phatthanasopon, P. Lertsuksakda, P. Kulsiri, T. Kerdsri, P. Pupatwibul, T. Bamrung, J. Piyawongwisal, R. Phipathananunth, and S. Suwajanakorn, "Typhoon: Thai large language models," *arXiv preprint arXiv:2312.13951*, 2023. [Online]. Available: https://arxiv.org/abs/2312.13951

[3] K. Pipatanakul, P. Manakul, N. Nitarach, W. Sirichotedumrong, S. Nonesung, T. Jaknamon, P. Pengpun, P. Taveekitworachai, A. Na-Thalang, S. Sripaisarnmongkol, K. Jirayoot, and K. Tharnpipitchai, "Typhoon 2: A family of open text and multimodal thai large language models," *arXiv preprint arXiv:2412.13702*, 2024. [Online]. Available: https://arxiv.org/abs/2412.13702

[4] S. Gupta, Y. Matsubara, A. Chadha, and A. Moschitti, "Cross-lingual knowledge distillation for answer sentence selection in low-resource languages," in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 14 078–14 092. [Online]. Available: https://aclanthology.org/2023.findings-acl.885/

[5] V. Goyle, P. Krishnaswamy, K. G. Ravikumar, U. Chattopadhyay, and K. Goyle, "Neural machine translation for low resource languages," 2023. [Online]. Available: https://arxiv.org/abs/2304.07869

[6] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020, pp. 8440–8451. [Online]. Available: https://aclanthology.org/2020.acl-main.747/

[7] G. Karamanolakis, D. Hsu, and L. Gravano, "Cross-lingual text classification with minimal resources by transferring a sparse teacher," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association

for Computational Linguistics, 2020, pp. 3604–3622. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.323/

[8] D. Mekala, V. Gangal, and J. Shang, "Coarse2fine: Fine-grained text classification on coarsely-grained annotated data," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 583–594. [Online]. Available: https://aclanthology.org/2021.emnlp-main.46/

[9] Y. Fei, L. Cui, S. Yang, W. Lam, Z. Lan, and S. Shi, "Enhancing grammatical error correction systems with explanations," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 7489–7501. [Online]. Available: https://aclanthology.org/2023.acl-long.413/

[10] M. R. Qorib, A. F. Aji, and H. T. Ng, "Efficient and interpretable grammatical error correction with mixture of experts," in *Findings of the Association for Computational Linguistics: EMNLP 2024*. Miami, Florida, USA: Association for Computational Linguistics, 2024, pp. 17 127–17 138. [Online]. Available: https://aclanthology.org/2024.findings-emnlp.997/

[11] B. Dong, J. Bai, T. Xu, and Y. Zhou, "Large language models in education: A systematic review," in *2024 6th International Conference on Computer Science and Technologies in Education (CSTE)*, 2024, pp. 131–134.

[12] B. Waluyo and S. Kusumastuti, "Generative ai in student english learning in thai higher education: More engagement, better outcomes?" *Social Sciences & Humanities Open*, vol. 10, p. 101146, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590291124003437

[13] K. W. H. Chan, C. Bryant, L. Nguyen, A. Caines, and Z. Yuan, "Grammatical error correction for code-switched sentences by learners of English," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA and ICCL, 2024, pp. 7926–7938. [Online]. Available: https://aclanthology.org/2024.lrec-main.698/

[14] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. H. Awadallah, R. W. White, D. Burger, and C. Wang, "Autogen: Enabling next-gen llm applications via multi-agent conversation," 2023. [Online]. Available: https://arxiv.org/abs/2308.08155

[15] O. Khattab, A. Singhvi, P. Maheshwari, Z. Zhang, K. Santhanam, S. Vardhamanan, S. Haq, A. Sharma, T. T. Joshi, H. Moazam, H. Miller, M. Zaharia, and C. Potts, "Dspy: Compiling declarative language model calls into self-improving pipelines," 2023. [Online]. Available: https://arxiv.org/abs/2310.03714

[16] P. Phatraphureeraks, *Perfect English Grammar: Thai Grammar Textbook for English Learning*. Bangkok, Thailand: MIS Book, 2020, thai language grammar textbook for English learners.

[17] W. H. Lee, "Nlp tense prediction," 2023, kaggle Notebook demonstrating NLP-based tense prediction techniques. [Online]. Available: https://www.kaggle.com/code/leewanhung/nlp-tense-prediction

[18] H. Rahman, A. Hossain, and M. S. Hossain, "English tense classification," 2024. [Online]. Available: https://www.kaggle.com/dsv/8693486

Appendices

Appendix A

Ethics Approval Documents

This appendix contains the complete ethics documentation for both the user survey study and the dataset annotation work conducted as part of this research project.

A.1 Survey Study Ethics Documentation

A.1.1 Participant Information Sheet

The following pages contain the participant information sheet (Version 3, July 3rd, 2025) that was provided to all survey participants.

University of Sheffield | School of Computer Science

Participant Information Sheet

(แบบฟอร์มข้อมูลสำหรับผู้เข้าร่วมงานวิจัย)

University of Sheffield

(มหาวิทยาลัยเชฟฟิลด์)

Department of Computer Science

(ภาควิชาวิทยาการคอมพิวเตอร์)

Research Project Title and Researchers

(ชื่อโครงการวิจัยและนักวิจัย)

Project Title *(ชื่อโครงการ)*: Context-Aware Thai-English Translation with Integrated Grammar Learning: An LLM-Based Educational Tool

(การแปลภาษาไทย-อังกฤษที่เข้าใจบริบทพร้อมการเรียนรู้ไวยากรณ์แบบบูรณาการ: เครื่องมือการศึกษาที่ใช้โมเดลภาษาขนาดใหญ่)

Lead Researcher *(นักวิจัยหลัก)*: Papatchol Thientong

(นาย ปภัสชล เทียนทอง)

Student Researcher, MSc Computer Science

(นักศึกษาวิจัย ปริญญาโทวิทยาการคอมพิวเตอร์)

Email *(อีเมล)*: [pthientong1@sheffield.ac.uk]

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Supervisor *(อาจารย์ที่ปรึกษา)*: Ms. Varvara Papazoglou

Email *(อีเมล)*: [v.papazoglou@sheffield.ac.uk]

Department of Computer Science, University of Sheffield

(ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชฟฟิลด์)

Invitation to Participate

(คำเชิญให้เข้าร่วมงานวิจัย)

We are inviting you to take part in a research project.

(เราขอเชิญท่านเข้าร่วมในโครงการวิจัยนี้)

Before you decide if you would like to take part, it is important for you to understand why the research is being done and what it will involve.

(ก่อนที่ท่านจะตัดสินใจว่าต้องการเข้าร่วมหรือไม่ สิ่งสำคัญคือท่านต้องเข้าใจว่าทำไมงานวิจัยนี้จึงถูกดำเนินการและจะเกี่ยวข้องกับอะไรบ้าง)

Please read the following information carefully and discuss it with others if you wish.

(กรุณาอ่านข้อมูลต่อไปนี้อย่างรอบคอบและปรึกษากับผู้อื่นหากท่านต้องการ)

Ask us if there is anything you don't understand or if you would like more information.

(กรุณาสอบถามเราหากมีสิ่งใดที่ท่านไม่เข้าใจหรือต้องการข้อมูลเพิ่มเติม)

Take time to decide whether or not you want to take part.

(ใช้เวลาในการตัดสินใจว่าท่านต้องการเข้าร่วมหรือไม่)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

What is the project's purpose?

(วัตถุประสงค์ของโครงการคืออะไร?)

This research project is being conducted as part of a Master's dissertation in Computer Science at the University of Sheffield.

(โครงการวิจัยนี้ดำเนินการเป็นส่วนหนึ่งของวิทยานิพนธ์ปริญญาโทในสาขาวิทยาการคอมพิวเตอร์ที่มหาวิทยาลัยเชฟฟิลด์)

The project aims to develop and evaluate an educational web-based tool that helps Thai English learners improve their grammar skills through context-aware translation.

(โครงการนี้มีเป้าหมายในการพัฒนาและประเมินเครื่องมือเว็บไซต์เพื่อการศึกษาที่ช่วยผู้เรียนภาษาอังกฤษชาวไทยปรับปรุงทักษะไวยากรณ์ผ่านการแปลที่เข้าใจบริบท)

The tool uses advanced language processing technology to translate Thai sentences into English while providing explanations about grammatical choices and patterns.

(เครื่องมือนี้ใช้เทคโนโลยีการประมวลผลภาษาขั้นสูงในการแปลประโยคภาษาไทยเป็นภาษาอังกฤษพร้อมทั้งให้คำอธิบายเกี่ยวกับการเลือกใช้ไวยากรณ์และรูปแบบต่างๆ)

This research will help us understand how technology can better support language learning and whether such tools are useful for Thai English learners.

(งานวิจัยนี้จะช่วยให้เราเข้าใจว่าเทคโนโลยีสามารถสนับสนุนการเรียนรู้ภาษาได้ดีขึ้นอย่างไร และเครื่องมือดังกล่าวมีประโยชน์สำหรับผู้เรียนภาษาอังกฤษชาวไทยหรือไม่)

The study will run from July 2025 to September 2025, and your participation will involve completing two online questionnaires and one tool testing session over a period of approximately 6 weeks, with a total time commitment of about 60 minutes.

(การศึกษานี้จะดำเนินการตั้งแต่เดือนกรกฎาคม 2568 ถึงเดือนกันยายน 2568 และการเข้าร่วมของท่านจะประกอบด้วยการตอบแบบสอบถามออนไลน์ 2 ชุดและการทดสอบเครื่องมือ 1 ครั้งในช่วงประมาณ 6 สัปดาห์ โดยใช้เวลารวมประมาณ 60 นาที)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Why have I been chosen?

(ทำไมฉันถึงถูกเลือก?)

You have been invited to participate because you are

(ท่านได้รับเชิญให้เข้าร่วมเพราะท่าน):

- A Thai speaker learning English as a second language

  (เป็นผู้พูดภาษาไทยที่เรียนภาษาอังกฤษเป็นภาษาที่สอง)

- 18 years of age or older

  (อายุ 18 ปีขึ้นไป)

- Able to participate in an English-language study

  (สามารถเข้าร่วมในการศึกษาที่ใช้ภาษาอังกฤษได้)

We are recruiting participants through two phases: first, a broad demographic survey of Thai English learners (approximately 50-100 participants), followed by a more focused evaluation with a smaller group (20-30 participants) who will test the developed tool.

(เรากำลังคัดเลือกผู้เข้าร่วมผ่านสองระยะ: แรกคือการสำรวจข้อมูลทั่วไปของผู้เรียนภาษาอังกฤษชาวไทยในวงกว้าง (ประมาณ 50-100 คน) ตามด้วยการประเมินที่เจาะจงมากขึ้นกับกลุ่มเล็กกว่า (20-30 คน) ที่จะทดสอบเครื่องมือที่พัฒนาขึ้น)

Do I have to take part? *(ฉันจำเป็นต้องเข้าร่วมหรือไม่?)*

Taking part in this research is completely voluntary.

(การเข้าร่วมในงานวิจัยนี้เป็นการสมัครใจโดยสมบูรณ์)

If you do not wish to take part, there will be no negative consequences.

(หากท่านไม่ต้องการเข้าร่วม จะไม่มีผลกระทบเชิงลบใด ๆ)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

If you do decide to take part, you will be given this information sheet to keep and will be asked to sign a consent form.

(หากท่านตัดสินใจเข้าร่วม ท่านจะได้รับแบบฟอร์มข้อมูลนี้ไว้เก็บและจะถูกขอให้ลงนามในแบบฟอร์มยินยอม)

You can still withdraw at any time without giving a reason and without any negative consequences.

(ท่านยังสามารถถอนตัวได้ตลอดเวลาโดยไม่ต้องให้เหตุผลและไม่มีผลกระทบเชิงลบใด ๆ)

If you wish to withdraw from the research, please contact

(หากท่านต้องการถอนตัวจากงานวิจัย กรุณาติดต่อ) [pthientong1@sheffield.ac.uk].

Please note that if you choose to withdraw, it may not be possible to remove your data if you withdraw after June 14, 2025, as data will have been anonymised and included in the analysis.

(โปรดทราบว่าหากท่านเลือกที่จะถอนตัว อาจไม่สามารถลบข้อมูลของท่านได้หากท่านถอนตัวหลังจากวันที่ 14 กรกฎาคม พ.ศ. 2568 เนื่องจากข้อมูลจะถูกทำให้ไม่ระบุตัวตนและรวมในการวิเคราะห์แล้ว)

Taking part in this research is voluntary and will not create a legally binding agreement or any form of employment with the University of Sheffield.

(การเข้าร่วมในงานวิจัยนี้เป็นการสมัครใจและจะไม่ก่อให้เกิดข้อตกลงที่มีผลผูกพันทางกฎหมายหรือการจ้างงานในรูปแบบใดๆ กับมหาวิทยาลัยเชฟฟิลด์)

What will happen to me if I take part? What do I have to do?

(จะเกิดอะไรขึ้นกับฉันหากฉันเข้าร่วม? ฉันต้องทำอะไร?)

Your participation may involve one or both phases of the research, depending on your interest and availability, with a total time commitment of 20-30 minutes and 60 minutes for 2 phases.

(การเข้าร่วมของท่านอาจเกี่ยวข้องกับหนึ่งหรือทั้งสองระยะของการวิจัย ขึ้นอยู่กับความสนใจและความพร้อมของท่าน โดยใช้เวลารวม 20-30 นาที และ 60 นาทีสำหรับทั้งสองเฟส)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

1. Phase 1: General English Learning Survey (July 2025) - approximately 20-30 minutes

   (ระยะที่ 1: การสำรวจการเรียนภาษาอังกฤษทั่วไป (กรกฎาคม 2568) - ประมาณ 20-30 นาที)

   You will receive a Google Form link via email to complete at your convenience.

   (ท่านจะได้รับลิงก์ Google Form ทางอีเมลเพื่อกรอกตามความสะดวกของท่าน)

   - Questions about your background and experience learning English

     (คำถามเกี่ยวกับภูมิหลังและประสบการณ์การเรียนภาษาอังกฤษของท่าน)

   - Your current English proficiency level

     (ระดับความสามารถในภาษาอังกฤษปัจจุบันของท่าน)

   - Challenges you face with English grammar

     (ความท้าทายที่ท่านพบเกี่ยวกับไวยากรณ์ภาษาอังกฤษ)

   - Previous experience with translation tools

     (ประสบการณ์ก่อนหน้านี้กับเครื่องมือแปลภาษา)

2. Phase 2: Tool Testing and Evaluation (August 2025) - approximately 30 minutes

   (ระยะที่ 2: การทดสอบและประเมินเครื่องมือ (สิงหาคม 2568) - ประมาณ 30 นาที)

   If you choose to participate in this phase, you will:

   (หากท่านเลือกเข้าร่วมในระยะนี้ ท่านจะ):

   - Test the Thai-English translation tool through a web interface

     (ทดสอบเครื่องมือแปลภาษาไทย-อังกฤษผ่านเว็บอินเตอร์เฟซ)

   - Complete a feedback questionnaire about the tool's effectiveness and usability

     (ตอบแบบสอบถามความคิดเห็นเกี่ยวกับประสิทธิภาพและการใช้งานของเครื่องมือ)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Participant Code Generation

(การสร้างรหัสผู้เข้าร่วม)

To protect your privacy while allowing you to participate in both phases if desired, you will generate your own unique participant code using a simple algorithm based on personal information that cannot be traced back to you:

(เพื่อปกป้องความเป็นส่วนตัวของท่านในขณะที่ยอมให้ท่านเข้าร่วมทั้งสองระยะหากต้องการ ท่านจะสร้างรหัสผู้เข้าร่วมเฉพาะตัวของท่านเองโดยใช้อัลกอริธึมง่าย ๆ ที่อิงจากข้อมูลส่วนบุคคลที่ไม่สามารถติดตามกลับมาถึงตัวท่านได้):

1. First letter of your birth month (e.g., January = J, March = M)

   (อักษรตัวแรกของเดือนที่เกิด (เช่น มกราคม = J, มีนาคม = M)

2. Second digit of your age (e.g., if 24 years old, use 4)

   (หลักที่สองของอายุ (เช่น หากอายุ 24 ปี ใช้ 4)

3. Number of letters in your first name modulo 10

   (จำนวนตัวอักษรในชื่อแรกในภาษาอังกฤษของท่าน หารด้วย 10 เอาเศษ (เช่นถ้าชื่อภาษาอังกฤษท่านมี 11 ตัว = 1)

4. Number of siblings modulo 10

   (จำนวนพี่น้องหารด้วย 10 เอาเศษ ถ้าไม่มี = 0)

5. Last digit of your phone number

   (หลักสุดท้ายของหมายเลขโทรศัพท์)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

What information will be collected?

(ข้อมูลอะไรจะถูกเก็บรวบรวม?)

- Only your age and your self-generated participant code

  (เฉพาะอายุของท่านและรหัสผู้เข้าร่วมที่ท่านสร้างเอง)

- Your responses to questionnaires

  (คำตอบของท่านในแบบสอบถาม)

- Feedback about tool usage (Phase 2 participants only)

  (ความคิดเห็นเกี่ยวกับการใช้เครื่องมือ (เฉพาะผู้เข้าร่วมระยะที่ 2)

Your responsibilities

(ความรับผิดชอบของท่าน)

- Generate and remember your participant code for consistent identification across phases

  (สร้างและจำรหัสผู้เข้าร่วมของท่านเพื่อการระบุตัวตนที่สอดคล้องกันในทุกระยะ)

- Complete questionnaires honestly and thoroughly

  (ตอบแบบสอบถามอย่างสุจริตและละเอียด)

- participating in Phase 2, test the tool and provide constructive feedback

  (หากเข้าร่วมระยะที่ 2 ให้ทดสอบเครื่องมือและให้ความคิดเห็นที่สร้างสรรค์)

There are no lifestyle restrictions or ongoing commitments required for this study.

(ไม่มีข้อจำกัดในการดำเนินชีวิตหรือภาระผูกพันต่อเนื่องที่จำเป็นสำหรับการศึกษานี้)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

What are the possible disadvantages and risks of taking part?

(ข้อเสียและความเสี่ยงที่เป็นไปได้ของการเข้าร่วมคืออะไร?)

The risks involved in this study are minimal.

(ความเสี่ยงที่เกี่ยวข้องในการศึกษานี้มีน้อยมาก)

You may experience *(ท่านอาจประสบ)*:

- Mild fatigue from concentrating during the tool testing session(s)

  (ความเหนื่อยล้าเล็กน้อยจากการมีสมาธิระหว่างเซสชันทดสอบเครื่องมือ)

- Potential frustration if the translation tool does not work as expected during testing (phase 2 only)

  (ความหงุดหงิดที่เป็นไปได้หากเครื่องมือแปลภาษาไม่ทำงานตามที่คาดหวังระหว่างการทดสอบ (เฉพาะระยะที่ 2))

- The need to complete questionnaires / interview at specific times (early July and mid-August)

  (ความจำเป็นในการตอบแบบสอบถามและสัมภาษณ์ในช่วงเวลาเฉพาะ (ต้นเดือนกรกฎาคมและกลางเดือนสิงหาคม))

If you experience any discomfort during the tool testing session, you can take breaks or withdraw at any time.

(หากท่านรู้สึกไม่สบายระหว่างเซสชันทดสอบเครื่องมือ ท่านสามารถหยุดพักหรือถอนตัวได้ตลอดเวลา)

All activities can be completed at your own pace within the specified timeframes.

(กิจกรรมทั้งหมดสามารถทำได้ตามจังหวะของท่านเองภายในกรอบเวลาที่กำหนด)

University of Sheffield | School of Computer Science

What are the possible benefits of taking part?

(ผลประโยชน์ที่เป็นไปได้ของการเข้าร่วมคืออะไร?)

While there are no immediate personal benefits for taking part in this project, it is hoped that this work will contribute to the development of better educational tools for Thai English learners.

(แม้ว่าจะไม่มีผลประโยชน์ส่วนบุคคลโดยตรงจากการเข้าร่วมในโครงการนี้ แต่หวังว่างานนี้จะช่วยในการพัฒนาเครื่องมือการศึกษาที่ดีกว่าสำหรับผู้เรียนภาษาอังกฤษชาวไทย)

During the tool testing session in phase 2, you may find the translation tool interesting to try, and some participants report that using such tools helps them think about English grammar in new ways.

(ระหว่างเซสชันทดสอบเครื่องมือในระยะที่สอง 2 ท่านอาจพบว่าเครื่องมือแปลภาษานี้น่าสนใจที่จะลอง และผู้เข้าร่วมบางคนรายงานว่าการใช้เครื่องมือดังกล่าวช่วยให้พวกเขาคิดเกี่ยวกับไวยากรณ์ภาษาอังกฤษในแบบใหม่)

Will my taking part in this project be kept confidential?

(การเข้าร่วมของฉันในโครงการนี้จะถูกเก็บเป็นความลับหรือไม่?)

All your personal information will be kept strictly confidential and will only be accessible to members of the research team.

(ข้อมูลส่วนบุคคลทั้งหมดของท่านจะถูกเก็บเป็นความลับอย่างเข้มงวดและจะสามารถเข้าถึงได้เฉพาะสมาชิกของทีมวิจัยเท่านั้น)

Your name will never be used in any reports or publications from this research.

(ชื่อของท่านจะไม่ถูกใช้ในรายงานหรือสิ่งพิมพ์ใดๆ จากงานวิจัยนี้เลย)

You will generate your own participant code that cannot be traced back to your identity, ensuring complete anonymity in all research data.

(ท่านจะสร้างรหัสผู้เข้าร่วมของท่านเองที่ไม่สามารถติดตามกลับมาถึงตัวตนของท่านได้ ทำให้มั่นใจได้ว่าข้อมูลการวิจัยทั้งหมดจะไม่ระบุชื่อโดยสมบูรณ์)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Contact details (if provided) will be kept separately and securely, and will be deleted immediately after the research is completed.

(รายละเอียดการติดต่อ (หากมีการให้ไว้) จะถูกเก็บแยกต่างหากและอย่างปลอดภัย และจะถูกลบทันทีหลังจากการวิจัยเสร็จสิ้น)

All data will be stored securely on University of Sheffield systems that meet data protection standards.

(ข้อมูลทั้งหมดจะถูกเก็บอย่างปลอดภัยในระบบของมหาวิทยาลัยเชฟฟิลด์ที่เป็นไปตามมาตรฐานการปกป้องข้อมูล)

Data Protection

(การปกป้องข้อมูล)

What data will be collected *(ข้อมูลอะไรจะถูกเก็บรวบรวม)*:

Only your age, self-generated participant code, and questionnaire responses.

(เฉพาะอายุของท่าน รหัสผู้เข้าร่วมที่ท่านสร้างเอง และคำตอบแบบสอบถาม)

Legal basis for processing *(ฐานทางกฎหมายสำหรับการประมวลผล)*:

This research is carried out 'in the public interest' in accordance with data protection legislation.

(งานวิจัยนี้ดำเนินการ 'เพื่อประโยชน์สาธารณะ' ตามกฎหมายการปกป้องข้อมูล)

Further information can be found in the University's Privacy Notice *(ข้อมูลเพิ่มเติมสามารถพบได้ในประกาศความเป็นส่วนตัวของมหาวิทยาลัย)*: https://www.sheffield.ac.uk/govern/data-protection/privacy/general

Data Controller *(ผู้ควบคุมข้อมูล)*:

The University of Sheffield will act as the Data Controller for this study.

(มหาวิทยาลัยเชฟฟิลด์จะทำหน้าที่เป็นผู้ควบคุมข้อมูลสำหรับการศึกษานี้)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

This means that the University is responsible for looking after your information and using it properly.

(นี่หมายความว่ามหาวิทยาลัยมีความรับผิดชอบในการดูแลข้อมูลของท่านและใช้งานอย่างเหมาะสม)

Data storage and retention *(การเก็บรักษาข้อมูล)*:

Your questionnaire responses and interview data (if applicable) will be stored securely on University systems for up to 5 years after the completion of the research, as required by University policy.

(คำตอบแบบสอบถามและข้อมูลการสัมภาษณ์ของท่าน (หากมี) จะถูกเก็บอย่างปลอดภัยในระบบมหาวิทยาลัยสูงสุด 5 ปีหลังจากการวิจัยเสร็จสิ้น ตามที่นโยบายมหาวิทยาลัยกำหนด)

After this time, the data will be securely deleted.

(หลังจากเวลานี้ ข้อมูลจะถูกลบอย่างปลอดภัย)

Data sharing *(การแบ่งปันข้อมูล)*:

Anonymized data from this study may be used in academic publications and presentations.

(ข้อมูลที่ไม่ระบุตัวตนจากการศึกษานี้อาจถูกใช้ในสิ่งพิมพ์ทางวิชาการและการนำเสนอ)

No individual participants will be identifiable in any outputs.

(ผู้เข้าร่วมรายบุคคลจะไม่สามารถระบุตัวตนได้ในผลลัพธ์ใด ๆ)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Your rights *(สิทธิของท่าน)*:

You have the right to access, correct, or request deletion of your personal data.

(ท่านมีสิทธิในการเข้าถึง แก้ไข หรือขอให้ลบข้อมูลส่วนบุคคลของท่าน)

For more information about your rights and how to exercise them, please see the University's Privacy Notice. *(สำหรับข้อมูลเพิ่มเติมเกี่ยวกับสิทธิของท่านและวิธีการใช้สิทธิ กรุณาดูประกาศความเป็นส่วนตัวของมหาวิทยาลัย)*

What happens to the results of the research?

(จะเกิดอะไรขึ้นกับผลการวิจัย?)

The results of this research will be written up as part of a Master's dissertation, which will be submitted in September 2025.

(ผลการวิจัยนี้จะถูกเขียนเป็นส่วนหนึ่งของวิทยานิพนธ์ปริญญาโท ซึ่งจะส่งในเดือนกันยายน 2568)

Key findings may also be presented at academic conferences or published in research journals focused on educational technology or language learning.

(ผลการค้นพบที่สำคัญอาจถูกนำเสนอในการประชุมทางวิชาการหรือตีพิมพ์ในวารสารวิจัยที่เน้นเทคโนโลยีการศึกษาหรือการเรียนรู้ภาษา)

You will not be identified in any report or publication.

(ท่านจะไม่ถูกระบุตัวตนในรายงานหรือสิ่งพิมพ์ใด ๆ)

If you would like to receive a summary of the research findings, please let us know and we can email you a brief summary once the project is complete.

(หากท่านต้องการรับสรุปผลการวิจัย กรุณาแจ้งให้เราทราบ และเราสามารถส่งสรุปย่อทางอีเมลให้ท่านเมื่อโครงการเสร็จสิ้น)

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Who has ethically reviewed the project?

(ใครได้ทำการทบทวนโครงการด้านจริยธรรม?)

This project has been ethically approved by the School of Computer Science at the University of Sheffield as part of the University's Ethics Review Procedure.

(โครงการนี้ได้รับการอนุมัติด้านจริยธรรมจากคณะวิทยาการคอมพิวเตอร์ที่มหาวิทยาลัยเชฟฟิลด์เป็นส่วนหนึ่งของขั้นตอนการทบทวนจริยธรรมของมหาวิทยาลัย)

What if something goes wrong and I wish to complain about the research or report a concern?

(จะเป็นอย่างไรหากมีสิ่งผิดพลาดเกิดขึ้นและฉันต้องการร้องเรียนเกี่ยวกับงานวิจัยหรือรายงานข้อกังวล?)

If you would like to complain about any aspect of the research or if you feel that you have been exploited, abused or harmed as a result of taking part in the research, please contact

(หากท่านต้องการร้องเรียนเกี่ยวกับด้านใดด้านหนึ่งของงานวิจัยหรือหากท่านรู้สึกว่าถูกเอาเปรียบ ทำร้าย หรือเป็นอันตรายจากการเข้าร่วมงานวิจัย กรุณาติดต่อ):

Primary Contact *(ผู้ติดต่อหลัก)*: Papatchol Thientong *(นาย ปภัสชล เทียนทอง)*

Email *(อีเมล)*: [pthientong1@sheffield.ac.uk]

Supervisor *(อาจารย์ที่ปรึกษา)*: Ms. Varvara Papazoglou

Email *(อีเมล)*: [v.papazoglou@sheffield.ac.uk]

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Independent Contact (ผู้ติดต่ออิสระ):

Principal Ethics Contact for the School of Computer Science (ผู้ติดต่อหลักด้านจริยธรรมของคณะวิทยาการคอมพิวเตอร์)

- Name *(ชื่อ)*: Prof. Jon Barker
- Email *(อีเมล)*: [j.p.barker@sheffield.ac.uk]
- Department *(แผนก)*: Computer Science *(วิทยาการคอมพิวเตอร์)*

Head of Department (หัวหน้าแผนก)

- Name *(ชื่อ)*: Professor Heidi Christensen
- Email *(อีเมล)*: [heidi.christensen@sheffield.ac.uk]

If your complaint is about how your personal data has been handled, you can find information about how to raise a complaint in the University's Privacy Notice

(หากข้อร้องเรียนของท่านเกี่ยวกับวิธีการจัดการข้อมูลส่วนบุคคลของท่าน ท่านสามารถหาข้อมูลเกี่ยวกับวิธีการยื่นข้อร้องเรียนได้ในประกาศความเป็นส่วนตัวของมหาวิทยาลัย): https://www.sheffield.ac.uk/govern/data-protection/privacy/general

You will be given a copy of this information sheet and a signed consent form to keep.

(ท่านจะได้รับสำเนาของแบบฟอร์มข้อมูลนี้และแบบฟอร์มยินยอมที่ลงนามแล้วไว้เก็บ)

Thank you for considering taking part in this research project. Your participation would be greatly valued and will contribute to improving educational tools for Thai English learners.

(ขอบคุณที่ได้พิจารณาเข้าร่วมในโครงการวิจัยนี้ การเข้าร่วมของท่านจะมีค่าอย่างมากและจะช่วยในการปรับปรุงเครื่องมือการศึกษาสำหรับผู้เรียนภาษาอังกฤษชาวไทย)

Version 3 – July 3rd, 2025

A.1.2 Participant Consent Form

The following pages contain the consent form (Version 3, July 3rd, 2025) that participants signed before taking part in the survey.

University of Sheffield | School of Computer Science

Context-Aware Thai-English Translation with Integrated Grammar Learning: An LLM-Based Educational Tool

(การแปลภาษาไทย-อังกฤษที่เข้าใจบริบทพร้อมการเรียนรู้ไวยากรณ์แบบบูรณาการ: เครื่องมือการศึกษาที่ใช้โมเดลภาษาขนาดใหญ่)

Participant Consent Form

(แบบฟอร์มยินยอมของผู้เข้าร่วม)

Please tick the appropriate boxes

(โปรดเลือกตัวเลือกที่เหมาะสม)

	Yes (ใช่)	No (ไม่ใช่)
Taking Part in the Project ***(การเข้าร่วมโครงการ)***		
I have read and understood the project information sheet dated July 3rd, 2025 or the project has been fully explained to me. (If you answer No to this question please do not fill out the rest of the form until you understand what taking part in the project will mean.) *(ข้าพเจ้าได้อ่านและเข้าใจเอกสารข้อมูลโครงการลงวันที่ 3 กรกฎาคม พ.ศ. 2568 หรือได้รับการอธิบายโครงการอย่างครบถ้วนแล้ว (หากคุณตอบ "ไม่ใช่" ต่อคำถามนี้ โปรดอย่ากรอกส่วนที่เหลือของแบบฟอร์มจนกว่าคุณจะเข้าใจว่าการเข้าร่วมโครงการหมายความว่าอย่างไร))*	☐	☐
I have been given the chance to ask questions about the project. *(ข้าพเจ้าได้รับโอกาสในการถามคำถามเกี่ยวกับโครงการ)*	☐	☐
I understand how to raise a concern or make a complaint. *(ข้าพเจ้าเข้าใจวิธีการแสดงความกังวลหรือการร้องเรียน)*	☐	☐

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

	Yes (ใช่)	No (ไม่ใช่)
I agree to participate in Phase 1: General English Learning Survey (approximately 20-30 minutes) *(ข้าพเจ้าตกลงเข้าร่วมระยะที่ 1: การสำรวจการเรียนภาษาอังกฤษทั่วไป (ประมาณ 20-30 นาที))*	☐	☐
I agree to participate in Phase 2: Tool Testing and Evaluation (approximately 20-30 minutes) *(ข้าพเจ้าตกลงเข้าร่วมระยะที่ 2: การทดสอบและประเมินเครื่องมือ (ประมาณ 20-30 นาที))*	☐	☐
I understand that I can choose to participate in one or both phases of the research *(ข้าพเจ้าเข้าใจว่าสามารถเลือกเข้าร่วมหนึ่งหรือทั้งสองระยะของการวิจัยได้)*	☐	☐
I understand that I will generate my own participant code to protect my anonymity while allowing participation across phases *(ข้าพเจ้าเข้าใจว่าจะสร้างรหัสผู้เข้าร่วมของตนเองเพื่อปกป้องการไม่ระบุตัวตนในขณะที่ยอมให้เข้าร่วมข้ามระยะต่างๆ)*	☐	☐
I understand that only my age and self-generated participant code will be collected as personal data for this study *(ข้าพเจ้าเข้าใจว่ามีเพียงอายุและรหัสผู้เข้าร่วมที่สร้างขึ้นเองเท่านั้นที่จะถูกเก็บรวบรวมเป็นข้อมูลส่วนบุคคลสำหรับการศึกษานี้)*	☐	☐
I understand that by choosing to participate as a volunteer in this research, this does not create a legally binding agreement nor is it intended to create an employment relationship with the University of Sheffield. *(ข้าพเจ้าเข้าใจว่าการเลือกที่จะเข้าร่วมเป็นอาสาสมัครในการวิจัยนี้ไม่ได้สร้างข้อตกลงที่มีผลผูกพันทางกฎหมายและไม่มีเจตนาที่จะสร้างความสัมพันธ์ในการจ้างงานกับมหาวิทยาลัยเชฟฟิลด์)*	☐	☐
I understand that I can withdraw from the research/study at any time. I understand that I do not have to give any reasons for this and there will be no negative consequences if I choose to withdraw. *(ข้าพเจ้าเข้าใจว่าสามารถถอนตัวจากการวิจัย/การศึกษาได้ตลอดเวลา ข้าพเจ้าเข้าใจว่าไม่จำเป็นต้องให้เหตุผลใด ๆ และจะไม่มีผลเสียหายใด ๆ หากเลือกที่จะถอนตัว)*	☐	☐
I understand that if I choose to withdraw, it may not be possible to remove the research data I have contributed if I withdraw after June 14, 2025. *(ข้าพเจ้าเข้าใจว่าหากเลือกที่จะถอนตัว อาจไม่สามารถลบข้อมูลการวิจัยที่ข้าพเจ้าได้มีส่วนร่วมได้หากถอนตัวหลังจากวันที่ 14 กรกฎาคม พ.ศ. 2568)*	☐	☐

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

	Yes (ใช่)	No (ไม่ใช่)
How my information will be used during and after the project **(วิธีการใช้ข้อมูลของข้าพเจ้าในระหว่างและหลังโครงการ)**		
I understand my personal details such as name, phone number, address and email address etc. will not be seen by anyone outside the project team. *(ข้าพเจ้าเข้าใจว่ารายละเอียดส่วนบุคคลของข้าพเจ้า เช่น ชื่อ หมายเลขโทรศัพท์ ที่อยู่ และที่อยู่อีเมล ฯลฯ จะไม่ถูกเห็นโดยบุคคลใดนอกเหนือจากทีมโครงการ)*	☐	☐
I understand that my responses to the questionnaire(s) may be quoted anonymously in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs. *(ข้าพเจ้าเข้าใจว่าคำตอบของข้าพเจ้าต่อแบบสอบถามทั้งสองชุดอาจถูกอ้างอิงแบบไม่เปิดเผยชื่อในสิ่งพิมพ์ รายงาน หน้าเว็บ และผลงานวิจัยอื่นๆ ข้าพเจ้าเข้าใจว่าชื่อของข้าพเจ้าจะไม่ถูกระบุในผลงานเหล่านี้)*	☐	☐
I understand that anonymized questionnaire data may be used for research analysis and publication. *(ข้าพเจ้าเข้าใจว่าข้อมูลแบบสอบถามที่ไม่ระบุชื่ออาจถูกใช้สำหรับการวิเคราะห์การวิจัยและการตีพิมพ์)*	☐	☐
I understand that no audio or video recordings will be made during this study. *(ข้าพเจ้าเข้าใจว่าจะไม่มีการบันทึกวิดีโอระหว่างการศึกษานี้)*	☐	☐

Participant Details *(รายละเอียดผู้เข้าร่วม)*:

- Name of participant [printed] *(ชื่อผู้เข้าร่วม [ตัวพิมพ์])*: _____
- Signature *(ลายเซ็น)*: _____
- Date *(วันที่)*: _____

Researcher Details (รายละเอียดนักวิจัย):

- Name of Researcher [printed] *(ชื่อนักวิจัย [ตัวพิมพ์])*: Papatchol Thientong
- Signature *(ลายเซ็น)*: _____
- Date *(วันที่)*: _____

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Project Contact Details for Further Information

(รายละเอียดการติดต่อโครงการสำหรับข้อมูลเพิ่มเติม):

Principal Investigator/Lead Researcher *(หัวหน้าโครงการ/นักวิจัยหลัก)*:

- Name *(ชื่อ)*: Papatchol Thientong
- Email *(อีเมล)*: [pthientong1@sheffield.ac.uk]
- Department *(แผนก)*: Computer Science *(วิทยาการคอมพิวเตอร์)*

Supervisor (อาจารย์ที่ปรึกษา):

- Name *(ชื่อ)*: Ms. Varvara Papazoglou
- Email *(อีเมล)*: [v.papazoglou@sheffield.ac.uk]
- Department *(แผนก)*: Computer Science *(วิทยาการคอมพิวเตอร์)*

For complaints or concerns about the project

(สำหรับการร้องเรียนหรือข้อกังวลเกี่ยวกับโครงการ)

Principal Ethics Contact for the School of Computer Science (ผู้ติดต่อหลักด้านจริยธรรมของคณะวิทยาการคอมพิวเตอร์)

- Name *(ชื่อ)*: Prof. Jon Barker
- Email *(อีเมล)*: [j.p.barker@sheffield.ac.uk]
- Department *(แผนก)*: Computer Science *(วิทยาการคอมพิวเตอร์)*

Head of Department (หัวหน้าแผนก)

- Name *(ชื่อ)*: Professor Heidi Christensen
- Email *(อีเมล)*: [heidi.christensen@sheffield.ac.uk]

Version 3 – July 3rd, 2025

University of Sheffield | School of Computer Science

Data Protection Information *(ข้อมูลการปกป้องข้อมูล)*:

This research is conducted in accordance with the University of Sheffield's Data Protection Policy and GDPR requirements. Your data will be stored securely on University systems and will only be accessible to the research team. For more information about how your data is used, please refer to the project information sheet.

(การวิจัยนี้ดำเนินการให้สอดคล้องกับนโยบายการปกป้องข้อมูลของมหาวิทยาลัยเชฟฟิลด์และข้อกำหนด GDPR ข้อมูลของคุณจะถูกเก็บไว้อย่างปลอดภัยในระบบของมหาวิทยาลัยและสามารถเข้าถึงได้โดยทีมวิจัยเท่านั้น สำหรับข้อมูลเพิ่มเติมเกี่ยวกับการใช้ข้อมูลของคุณ โปรดดูแผ่นข้อมูลโครงการ)

Version 3 – July 3rd, 2025

A.1.3 Ethics Application for Survey Study

The following pages contain the formal ethics application submitted to the University of Sheffield Ethics Committee.

The University Of Sheffield.

Application 069760

Section A: Applicant details

Date application started:
Tue 17 June 2025 at 14:49

First name:
Papatchol

Last name:
Thientong

Email:
pthientong1@sheffield.ac.uk

Programme name:
Computer Science with Speech and Natural Language Processing

Module name:
COM Msc Dissertation Project
Last updated:
11/07/2025

Department:
Computer Science

Applying as:
Postgraduate Taught

Research project title:
Context-Aware Thai-English Translation with Integrated Grammar Learning: An LLM-Based Educational Tool

Has your research project undergone academic review, in accordance with the appropriate process?
Yes

Similar applications:
- not entered -

Section B: Basic information

Supervisor

Name	Email
Varvara Papazoglou	v.papazoglou@sheffield.ac.uk

Proposed project duration

Start date (of data collection):
Thu 10 July 2025

Anticipated end date (of project)
Mon 8 September 2025

3: Project code (where applicable)

Project externally funded?
No

Appendix A. Ethics Approval Documents

Project code
- not entered -

Suitability

Takes place outside UK?
Yes

Involves NHS?
No

Health and/or social care human-interventional study?
No

ESRC funded?
No

Likely to lead to publication in a peer-reviewed journal?
Yes

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial or a medical device study?
No

Involves social care services provided by a local authority?
No

Is social care research requiring review via the University Research Ethics Procedure
No

Involves adults who lack the capacity to consent?
No

Involves research on groups that are on the Home Office list of 'Proscribed terrorist groups or organisations?
No

Indicators of risk

Involves potentially vulnerable participants?
No
Involves potentially highly sensitive topics?
No

Section C: Summary of research

1. Aims & Objectives

This research aims to develop and evaluate a web-based educational tool that helps Thai students improve their English grammar skills through intelligent translation combined with explicit grammatical instruction. Thai learners face significant challenges with English grammar due to structural differences between the languages, particularly with verb tenses, articles, and word order patterns that do not exist in Thai. Current translation tools often correct mistakes without explaining the underlying grammatical rules, creating dependency without facilitating actual learning.

The research involves developing a hybrid artificial intelligence system using two specialized language models to translate Thai sentences into grammatically correct English while providing clear explanations of grammatical choices made during translation. The technical development will be followed by creation of a user-friendly web interface that presents both translations and educational explanations in Thai to maximize comprehension for learners.
The evaluation phase will involve Thai English learners testing the developed tool and providing feedback through structured questionnaires about its educational effectiveness and usability. The study will assess whether this approach of combining translation with explicit grammar instruction helps learners understand English grammatical concepts better than traditional correction-only methods. The

expected outcome is an evidence-based educational platform that serves as both a practical translation aid and an effective grammar learning tool, potentially improving English language education for Thai students while advancing our understanding of how artificial intelligence can be applied to support second language acquisition.

2. Methodology

This research will employ a mixed-methods approach combining technical development with user evaluation to assess the effectiveness of the educational translation tool. The study will be conducted entirely online to accommodate participants' schedules and enable broader participation from Thai English learners across different locations and backgrounds.

The research will be structured in two main phases, each involving online questionnaires administered through Google Forms. The first phase will target a broad demographic of Thai English learners to understand general learning patterns and needs. The second phase will focus on a more specific group who will directly test and evaluate the developed tool.

Phase 1: Initial User Needs Assessment
Participants will complete an online questionnaire exploring their current English learning methods, specific grammar challenges they face, and existing tools they use for language learning. This baseline survey will inform the tool's design and help identify the most common grammatical difficulties among Thai English learners. The questionnaire will take approximately 20-30 minutes to complete and will be distributed through university networks, English learning communities, and social media platforms to reach the widest possible range of Thai English learners. This phase aims to gather insights from diverse backgrounds including students, working professionals, and language center learners.

Phase 2: Tool Testing and Evaluation
Following the initial assessment and subsequent tool development using high-performance computing resources, a more specific group of participants will be recruited to directly test the web-based translation tool. These participants will use the tool by translating sample Thai sentences and reviewing the grammatical explanations provided. The testing session will be self-paced and conducted through a web browser, allowing participants to explore the tool's features at their own convenience. After completing the testing session, participants will complete a detailed feedback questionnaire evaluating the tool's usability, accuracy, translation quality, and educational value. This phase will assess whether the tool effectively explains grammar concepts and helps users understand English sentence structure.

Data Analysis
Quantitative data from both questionnaires will be analyzed statistically to identify patterns in learning preferences, demographic trends, and tool effectiveness ratings. Qualitative data from open-ended survey responses will be analyzed thematically to understand user experiences and identify areas for tool improvement. All participant responses will be anonymized using assigned participant codes, ensuring that individual users cannot be identified in the analysis or reporting of results.

3. Personal Safety

Have you completed your departmental risk assessment procedures, if appropriate?

Not Applicable

Raises personal safety issues?

No

This research poses no personal safety or physical/mental well-being risks because all activities will be conducted entirely online through secure platforms, focusing on the non-sensitive topic of language learning tool evaluation. The methodology involves standard academic procedures including voluntary questionnaires and interviews about educational technology, with no physical intervention, sensitive personal information, or potentially distressing content. Participants can withdraw at any time, and as the researcher, I will be working in secure university facilities with supervisor support, following established practices for educational technology research that are routinely conducted without safety concerns in academic settings.

Section D: About the participants

1. Potential Participants

Potential participants will be identified and recruited through established networks to ensure access to appropriate Thai English learners while maintaining ethical recruitment practices. The study targets adult Thai speakers who are actively engaged in English language learning, representing a population that can provide meaningful insights into the educational tool's effectiveness.

Primary Recruitment Sources:
The recruitment strategy focuses on two main channels that provide access to diverse groups of Thai English learners. Staff members from a tutoring school in Bangkok will be approached as they represent adults working in educational environments with regular English language exposure and practical experience with language learning tools. Additionally, participants will be recruited through university networks in Thailand, including students and recent graduates who are actively learning or using English in academic contexts. These sources ensure access to participants with varying levels of English proficiency and different educational backgrounds.

Selection Criteria:
Participants must be native Thai speakers aged 18 years and above who are currently learning or actively using English as a second language. They should possess sufficient English proficiency to understand research materials and provide meaningful feedback about the educational tool. Basic computer or smartphone skills are required for online participation, along with reliable internet access. All participants must demonstrate willingness to engage in online questionnaires and independent tool testing sessions.

Recruitment Process:
Recruitment will be conducted through direct contact via existing personal and professional networks, ensuring that potential participants receive clear explanations about the study purpose, procedures, and time commitments involved. The voluntary nature of participation will be emphasized throughout all recruitment communications, with no pressure or incentives offered. The target sample size of 20-30 participants will be recruited across different age groups and English proficiency levels to ensure diverse perspectives on the tool's educational effectiveness.

Ethical Considerations:
All recruitment activities will be conducted through pre-existing relationships and networks, eliminating concerns about coercion or undue influence. Participants will receive comprehensive information about the study enabling them to make fully informed decisions about participation. No academic, professional, or personal pressure will be applied during recruitment, and the voluntary nature of participation will be clearly communicated at all stages of the research process.

2. Recruiting Potential Participants

Potential participants will be approached through two main channels using a direct, informal recruitment process:

Tutoring School Staff (Bangkok):
Adult staff members at a tutoring school in Bangkok will be approached directly by the school administrator or through existing professional relationships. Bilingual recruitment materials (Thai and English) will be shared with interested staff members who are native Thai speakers actively learning English. The recruitment package will include information explaining that the research involves developing a web-based tool to help Thai learners improve their English grammar through translation, detailing participation requirements (completing two online questionnaires and independently testing an educational translation tool, with approximately 45 minutes total commitment over 6 weeks), and emphasizing that participation is completely voluntary with no professional consequences and the ability to withdraw at any time.

University Network:
Thai university students and graduates will be approached through my sister's university network via personal referrals and social connections. Recruitment materials will be shared through informal channels such as social media groups, personal messages, and word-of-mouth recommendations. Potential participants will receive the same bilingual information package explaining the research purpose, time commitment, and voluntary nature of participation.

Recruitment Process:
Those interested in learning more will contact the researcher directly via email or through the referring contact. This initial contact indicates only willingness to receive additional information and creates no obligation to participate. Interested individuals will then receive the full participant information sheet and consent form, allowing them to make an informed decision about participation. The recruitment target is 20-30 adult participants aged 18+ with varied English proficiency levels.
All recruitment emphasizes that participation is entirely voluntary, separate from any academic or professional obligations, and that participants can withdraw at any time without consequences.

2.1. Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by IT Services? No

- not entered -

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

Informed consent will be obtained through a comprehensive digital process where adult participants aged 18 years and above receive detailed bilingual information materials in Thai and English explaining the research purpose of developing an educational Thai-English translation tool, participation requirements including two online questionnaires and web tool testing totaling approximately 50-60 minutes over 6 weeks, and GDPR-compliant data handling procedures including the legal basis for processing (public interest research), data storage on secure University systems, access limitations to the research team only, and 5-year retention periods. Participants will provide granular digital consent for each research activity through online forms, with consent reconfirmed at each stage and clear emphasis on voluntary participation unrelated to any academic, professional, or personal obligations, along with contact information for the research team and independent complaint procedures to ensure participants can raise concerns at any time during the study.

4. Payment

Will financial/in kind payments be offered to participants? No

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

The potential for harm in this research is minimal, with the primary risks being minor psychological discomfort and inconvenience to participants. The most likely source of harm is temporary frustration or mild anxiety that participants might experience when testing the translation tool if it produces incorrect translations or confusing grammar explanations, potentially causing brief discouragement about their English learning abilities or feelings of inadequacy regarding their language skills. Additionally, participants might experience slight self-consciousness when completing questionnaires about their English proficiency levels, particularly given cultural tendencies in Thai educational contexts where students may feel modest or hesitant about discussing their language abilities. There is also minimal risk of data-related concern if participants worry about how their responses might be used despite anonymization procedures, though comprehensive information provision should mitigate this risk. The degree of potential harm is limited to minor inconvenience from the time commitment of approximately 50-60 minutes across the study period and temporary mild frustration with tool performance, which is justified by the research objectives of developing educational tools that could significantly benefit Thai English learners' grammar acquisition and contribute to improved language learning resources for this underserved population, with the potential benefits far outweighing the minimal risks involved.

How will this be managed to ensure appropriate protection and well-being of the participants?

Multiple safeguards will be implemented to minimize potential harm and ensure participant well-being throughout the research process. Clear expectations will be established about the experimental nature of the translation tool, with explicit explanations that errors and imperfections are expected as part of the development process, thereby reducing frustration when mistakes occur and emphasizing that tool performance does not reflect participants' English abilities. Participants will be reminded that their feedback contributes to improving educational resources rather than serving as an assessment of their language skills. Cultural sensitivity will be maintained by acknowledging Thai communication styles and providing materials in both Thai and English to ensure full comprehension, while allowing participants to express concerns or feedback in their preferred language. Participants will be explicitly informed of their right to skip questions, take breaks during testing sessions, pause questionnaires, or withdraw from the study at any time without providing explanations or facing any consequences. Contact information for both the primary researcher and university supervisor will be readily available for participants to raise concerns about any aspect of their involvement, and technical support will be provided promptly to address any difficulties with the web-based tool that might cause frustration. The research design incorporates flexibility in timing and pacing, allowing participants to complete activities at their own convenience to reduce any sense of pressure or time-related stress, while regular reminders about the voluntary nature of participation ensure ongoing comfort with their involvement in the study.

6. Potential harm to others who may be affected by the research activities

Which other people, if any, may be affected by the research activities, beyond the participants and the research team?

Other people who may be affected by this research include colleagues and administrators at the Bangkok tutoring school where recruitment will take place, friends and acquaintances within university networks who may be approached for participation or become aware of the research through social connections, and household members such as family or flatmates who may be present in participants' homes during online questionnaire completion or tool testing sessions. School staff who are not participating may become aware of the research through workplace discussions or recruitment activities, while university network contacts may encounter recruitment communications or hear about the study from potential participants. Household members represent the most direct form of indirect involvement as they may be present during participants' engagement with research activities, potentially overhearing responses to questionnaires about English learning experiences or observing participants' interactions with the translation tool during testing sessions.

What is the potential for harm to these people?

The potential harm to individuals beyond direct participants is minimal and primarily involves minor inconvenience or privacy considerations. School administrators may experience minimal additional workload in facilitating initial recruitment communication with interested staff members, though this represents only a brief administrative task rather than ongoing burden. Friends and acquaintances within university networks may experience minor inconvenience if asked about the research or if recruitment information is shared through social channels, potentially creating slight social pressure or awkwardness around declining to participate or share information. Household members such as family, flatmates, or others present during online research activities face the primary risk of inadvertently overhearing questionnaire responses or observing screen content, which could compromise participant privacy and lead to minor embarrassment if personal information about English learning challenges or experiences is overheard. There is also minimal risk that household members might feel obligated to provide privacy or quiet space for participants during research activities, potentially causing minor disruption to household routines or creating temporary inconvenience in shared living spaces. Overall, the potential for harm to these individuals remains very low, consisting primarily of brief administrative tasks, minor social interactions, and temporary privacy considerations that do not pose significant physical, psychological, or social risks to anyone beyond the direct research participants.

How will this be managed to ensure appropriate safeguarding of these people?

Several measures will be implemented to safeguard all affected parties while maintaining appropriate boundaries and minimizing any potential impact on non-participants. School administrators will receive clear information about the research scope and their minimal involvement, which is limited to facilitating initial contact with interested staff members through appropriate professional channels, with explicit assurance that no additional administrative burden will be placed on school operations and that staff participation remains entirely separate from professional duties. Recruitment through university networks will be conducted respectfully through existing social connections with clear messaging that participation is entirely voluntary and independent, ensuring no pressure is placed on mutual

friends or acquaintances to participate or facilitate recruitment activities. To protect household privacy, participants will receive comprehensive guidance about maintaining confidentiality during research activities, including recommendations to complete questionnaires and tool testing in private settings when possible, advice to use headphones if needed to prevent others from overhearing responses, and flexible scheduling options that accommodate participants' preferences for privacy and minimize household disruption. Professional and social boundaries will be maintained throughout all recruitment and research activities, with participants clearly informed that their decision to participate or withdraw will have no impact on existing professional or social relationships, while contact information for the research team will be provided to address any questions or concerns from participants or their networks. These comprehensive safeguards ensure minimal impact on non-participants while protecting participant privacy and maintaining appropriate boundaries between research activities and existing professional or social relationships.

7. Reporting of safeguarding concerns or incidents

What arrangements will be in place for participants, and any other people external to the University who are involved in, or affected by, the research, to enable reporting of incidents or concerns?

If participants or any other people are dissatisfied with any aspect of the research and wish to make a complaint, they can contact
the researcher (Papatchol Thientong: pthientong1@sheffield.ac.uk) or
the project supervisor (Varvara Papazoglou: v.papazoglou@sheffield.ac.uk) or
the Principal Ethics Contact (Prof. Jon Barker: j.p.barker@sheffield.ac.uk) in the first instance.

If they feel their complaint has not been handled in a satisfactory way then they can contact the Head of the School of Computer Science, Prof. Heidi Christensen: heidi.christensen@sheffield.ac.uk.

If the complaint concerns how their personal data has been handled, they can find information about how to raise a complaint in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy. This information is also included in the information sheet.

Who will be the Designated Safeguarding Contact(s)?

The researchers involved, the Principal Ethics Contact and the Head of the School of Computer Science, as specified in the previous question.

How will reported incidents or concerns be handled and escalated?

The DSC (Designated Safeguarding Contact) should try to obtain as much information as possible regarding the reported incidents or concerns, whilst also acting sensitively and providing reassurance regarding the process that will be undertaken for handling the report. The Research Ethics & Integrity Manager in Research Services will also be notified so that advice can be sought regarding the appropriate next steps.

If necessary, advice will be sought from one of more the following, depending on the nature of the concern: |
- The University's Safeguarding and Welfare Unit
- Human Resources - University Research Ethics Committee
- Vice-President for Research and/or Faculty Directors of Research & Innovation

The victim, and other relevant parties should be kept informed regarding progress and key decisions in dealing with the matter. (text adapted from https://www.sheffield.ac.uk/rs/ethicsandintegrity/safeguarding)

Section E: Personal data

1. Use of personal data

Will any personal data be processed or accessed as part of the project?
Yes

Will any 'special category' personal data be processed or accessed as part of the project?
No

Provide the number of people whose personal data you expect to process or access.
100

2. Managing personal data

Which organisation(s) will act as data controller(s) of the personal data?
University of Sheffield only

Who will have access to the personal data?
Access to personal data will be strictly limited to authorized personnel with legitimate research purposes and appropriate data protection training. Within the University of Sheffield, access will be restricted to myself as the primary researcher for data collection, analysis, and reporting purposes,

and my academic supervisor for research oversight and quality assurance. The Computer Science department's technical support staff may have limited access to anonymized data stored on university-approved systems for backup and technical maintenance purposes only, with no access to identifying information or personal details. External access will be minimal and carefully controlled, with the tutoring school administrator in Thailand having access only to basic contact information necessary for facilitating initial recruitment communication with interested adult staff members, but no access to research responses or collected data. University network contacts will only have access to basic study information for recruitment purposes, with no access to participant responses or research data. No other external collaborating organizations, partner institutions, or service providers will have access to personal data, and all individuals with data access will receive appropriate data protection guidance while data remains stored on University of Sheffield secure systems with access controls ensuring individuals can only access information necessary for their specific responsibilities.

What measures, processes and/or agreements will be put in place to manage the personal data?
Comprehensive data protection measures will be implemented throughout the research lifecycle to ensure participant privacy and confidentiality. Personal identifiers will be removed from all research data through a robust anonymization process where participants will generate their own unique alphanumeric codes using a predetermined algorithm that combines non-identifying elements including the first letter of their birth month, the second digit of their age, their first name length modulo 10, number of siblings modulo 10, and the last digit of their phone number (e.g., J-3-5-2-8). This self-generation system eliminates the need for a master list linking participant identities to codes, as participants can recreate their codes when needed while ensuring the algorithm components are sufficiently overlapping and non-unique to prevent reverse identification of individuals. All questionnaire responses will be processed using only these participant-generated codes, ensuring that research data cannot be directly linked to individual identities. Digital data management will follow strict security protocols with all data stored on University of Sheffield approved Google Drive systems with encryption, multi-factor authentication, and regular automated backups. Questionnaire responses will be collected through secure online platforms with SSL encryption, and consent forms will be scanned and stored digitally on secure university Google Drive before secure destruction of paper copies. Access limitations will be enforced through role-based permissions ensuring only authorized personnel can access specific data types, with no requirement for storing participant contact details alongside research responses since the self-generated code system eliminates the need for identity-linking mechanisms. Publication and dissemination safeguards will ensure participants cannot be identified in any reports, presentations, or academic publications through careful review of all outputs before release. Demographic information will be reported only in aggregate form, and any direct quotes from questionnaire responses will be thoroughly anonymized with identifying details removed or altered. Data retention policies will follow university guidelines with personal identifiers destroyed immediately after research completion in September 2025, while anonymized research data will be retained for 5 years as per university policy. Participants will be clearly informed about confidentiality measures and the self-generated code system, and provided with clear procedures for exercising their data subject rights including access, correction, and deletion of their personal information.

Will all identifiable personal data in digital or physical format be destroyed within a defined period after the project has ended?
Yes

When will the identifiable personal data be destroyed?
Different types of identifiable personal data will be destroyed according to appropriate timeframes based on their necessity for the research process and regulatory requirements: Participant Contact Details: Contact information including email addresses or other provided means of contact such as social media accounts, messaging platforms, or phone numbers used for sending questionnaire links and tool access will be destroyed immediately after completion of data collection (by September 2025), once all research activities have been completed and there is no further need to contact participants about the study. Digital Consent Forms: Signed consent forms will be retained for 5 years after dissertation submission in accordance with University of Sheffield research data retention policies, as these may be required for audit purposes or examination queries about research ethics compliance and will be destroyed by September 2030. Recruitment Materials: Any recruitment materials containing contact information will be securely deleted within 1 month after completion of the recruitment phase (by August 2025), once all necessary participants have been enrolled and no further recruitment activities are required. Anonymized Research Data: All anonymized research data including questionnaire responses using only participant-generated codes will be retained for 5 years after dissertation completion for potential future academic publications, with no identifiable personal information remaining after the destruction timeframes outlined above. Since participants generate their own codes using the predetermined algorithm, no master linking list exists to connect codes to identities, ensuring complete anonymization from the point of data collection. These timeframes balance research needs with data protection principles, ensuring personal data is not retained longer than necessary while allowing for appropriate academic review and potential future research use of fully anonymized data that cannot be traced back to individual participants.

3. Third-party services

Will any external third-party services not provided by the University be used to process or access personal data during the project?
No

4. Security of computers, devices and software

Will personal data be processed or accessed on any computers or devices that are not managed by the University of Sheffield?
Yes

Will all computers and devices that are not managed by the University of Sheffield be secured in accordance with the IT Code of Connection?
Yes

Will any software not approved by the University of Sheffield be used to process or access data?
No

Will any software be written or developed in order to process or access the personal data?
Yes

Will the new software be secured in accordance with the IT Code of Connection?
Yes

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project?
Yes

Document 1156603 (Version 4) — All versions
This information sheet is bilingual (Thai-English) to help with Thai participants.

Consent forms relevant to project?
Yes

Document 1156604 (Version 3) — All versions
This consent form is bilingual (Thai-English) to help with Thai participants.

Additional Documentation

Document 1156754 (Version 1) — All versions
Draft of the first questionnaire on Google Forms

External Documentation

- not entered -

Section G: Declaration

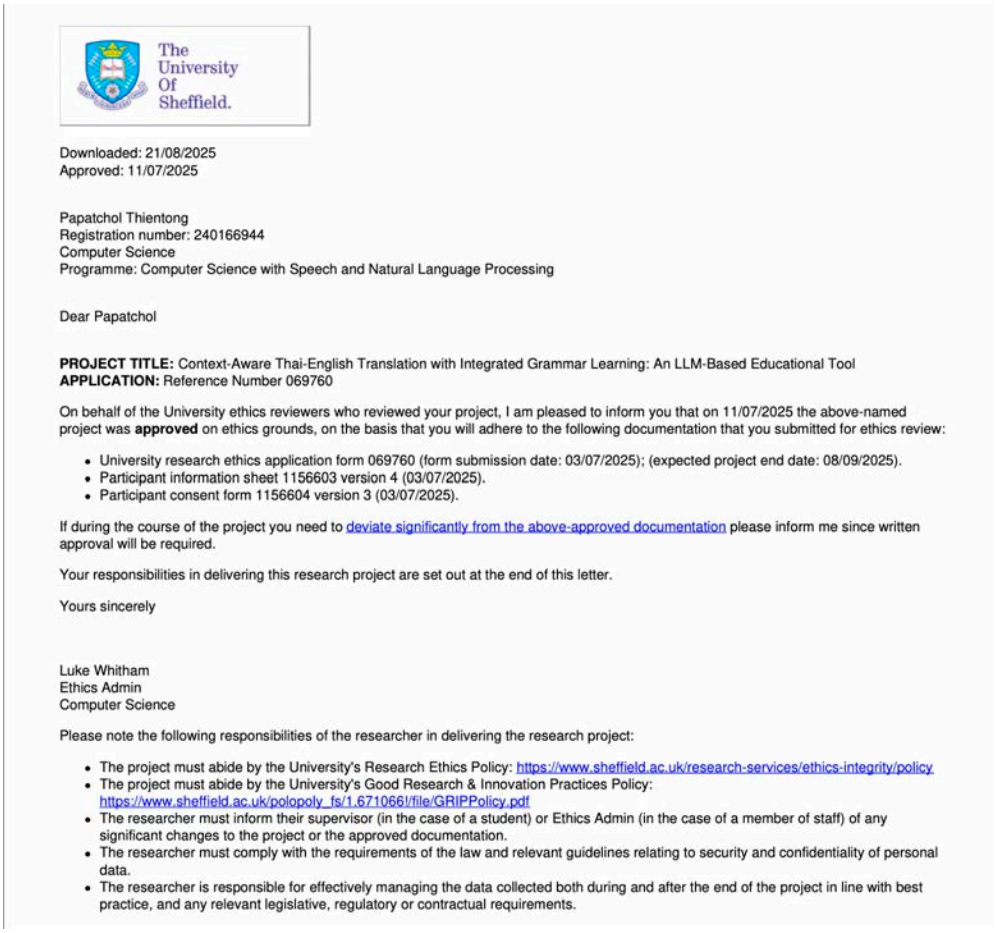
Signed by:
Papatchol Thientong
Date signed:
Thu 3 July 2025 at 15:36

Official notes

- not entered -

A.1.4 Ethics Approval Letter for Survey Study

The following document contains the official approval from the University Ethics Committee for the survey study.



The University Of Sheffield.

Downloaded: 21/08/2025
Approved: 11/07/2025

Papatchol Thientong
Registration number: 240166944
Computer Science
Programme: Computer Science with Speech and Natural Language Processing

Dear Papatchol

PROJECT TITLE: Context-Aware Thai-English Translation with Integrated Grammar Learning: An LLM-Based Educational Tool
APPLICATION: Reference Number 069760

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 11/07/2025 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 069760 (form submission date: 03/07/2025); (expected project end date: 08/09/2025).
- Participant information sheet 1156603 version 4 (03/07/2025).
- Participant consent form 1156604 version 3 (03/07/2025).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Luke Whitham
Ethics Admin
Computer Science

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Admin (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

A.2 Dataset Annotation Ethics Documentation

A.2.1 Ethics Application for Dataset Annotation

The following pages contain the ethics application for the dataset annotation work conducted to create training data for the XLM-RoBERTa tense classifier.

Application 070793

Section A: Applicant details

Date application started:
Mon 18 August 2025 at 15:55

First name:
Papatchol

Last name:
Thientong

Email:
pthientong1@sheffield.ac.uk

Programme name:
MSc Computer Science with Speech and Language Processing

Module name:
COM MSc Dissertation Project
Last updated:
20/08/2025

Department:
Computer Science

Applying as:
Postgraduate Taught

Research project title:
Thaislate: an AI-Enhanced Tool for Thai-English Grammar Learning

Has your research project undergone academic review, in accordance with the appropriate process?
Yes

Similar applications:
- not entered -

Section B: Basic information

Supervisor

Name	Email
Varvara Papazoglou	v.papazoglou@sheffield.ac.uk

Proposed project duration

Start date (of data collection):
Mon 18 August 2025

Anticipated end date (of project)
Mon 25 August 2025

3: Project code (where applicable)

Project externally funded?
No

Appendix A. Ethics Approval Documents

Project code
- not entered -

Suitability

Takes place outside UK?
No

Involves NHS?
No

Health and/or social care human-interventional study?
No

ESRC funded?
No

Likely to lead to publication in a peer-reviewed journal?
Yes

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial or a medical device study?
No

Involves social care services provided by a local authority?
No

Is social care research requiring review via the University Research Ethics Procedure
No

Involves adults who lack the capacity to consent?
No

Involves research on groups that are on the Home Office list of 'Proscribed terrorist groups or organisations?
No

Indicators of risk

Involves potentially vulnerable participants?
No
Involves potentially highly sensitive topics?
No

Section C: Summary of research

1. Aims & Objectives

The purpose of this research is to extend a tense prediction dataset, in order to create a system for tense classification, as a part of a larger system designed to help Thai students improve their English grammar skills.

2. Methodology

The existing dataset, retrieved from Kaggle [1,2], contains 13,316 English sentences, labelled with their grammatical tense as Past, Present, Future. I am planning to manually annotate a fraction (approximately 2,000 sentences) of this dataset with 24 fine-grained labels, specified in [3].

Citations:
[1] W. H. Lee, "NLP Tense Prediction," 2023, Kaggle Notebook demonstrating NLP-based tense prediction techniques. [Online]. Available: https://www.kaggle.com/code/leewanhung/ nlp-tense-prediction
[2] H. Rahman, A. Hossain, and M. S. Hossain, "English tense classification," 2024. [Online]. Available: https://www.kaggle.com/dsv/8693486

[3] P. Phatraphureeraks, Perfect English Grammar: Thai Grammar Textbook for English Learning. Bangkok, Thailand: MIS Book, 2020, Thai language grammar textbook for English learners.

3. Personal Safety

Have you completed your departmental risk assessment procedures, if appropriate?

Not Applicable

Raises personal safety issues?

No

No personal safety or physical or mental well-being concern is identified for the project.
As the only annotator, I will make sure to take regular breaks during the annotation process and will keep my supervisor up-to-date about the progress.

Section D: About the participants

1. Potential Participants

I am the only participant.

2. Recruiting Potential Participants

Not applicable.

2.1. Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by IT Services? No

- not entered -

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) No

I am the only participant, therefore I am aware of the details of the research.

4. Payment

Will financial/in kind payments be offered to participants? No

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

Physical tiredness and eye-strain are some of the potential harms, but are considered minor, and I will try to minimise their impact by taking regular breaks during the annotation process.

How will this be managed to ensure appropriate protection and well-being of the participants?

Reasonable, regular breaks will be taken and my supervisor will be kept up-to-date about the progress and potential issues or concerns.

6. Potential harm to others who may be affected by the research activities

Which other people, if any, may be affected by the research activities, beyond the participants and the research team?

No other people are involved in this project.

What is the potential for harm to these people?

Not applicable.

How will this be managed to ensure appropriate safeguarding of these people?

Not applicable.

7. Reporting of safeguarding concerns or incidents

What arrangements will be in place for participants, and any other people external to the University who are involved in, or affected by, the

research, to enable reporting of incidents or concerns?

I am the only participant, so in case of any problems or concerns I will inform my project supervisor (Varvara Papazoglou: v.papazoglou@sheffield.ac.uk) or the teaching support team of my academic department (com-teaching@sheffield.ac.uk) or my academic tutor (Nikolaos Aletras: n.aletras@sheffield.ac.uk).

Who will be the Designated Safeguarding Contact(s)?

My project supervisor, the teaching support team and my academic tutor.

How will reported incidents or concerns be handled and escalated?

Not applicable.

Section E: Personal data

1. Use of personal data

Will any personal data be processed or accessed as part of the project?
No

Are you sure that no personal data will be processed or accessed during your project?
Yes

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project?
No

Consent forms relevant to project?
No

Additional Documentation

External Documentation

- not entered -

Section G: Declaration

Signed by:
Papatchol Thientong
Date signed:
Mon 18 August 2025 at 16:09

Official notes

- not entered -

A.2.2 Ethics Approval Letter for Dataset Annotation

The following document contains the official approval from the University Ethics Committee for the dataset annotation work.

The University Of Sheffield.

Downloaded: 21/08/2025
Approved: 20/08/2025

Papatchol Thientong
Registration number: 240166944
Computer Science
Programme: MSc Computer Science with Speech and Language Processing

Dear Papatchol

PROJECT TITLE: Thaislate: an AI-Enhanced Tool for Thai-English Grammar Learning
APPLICATION: Reference Number 070793

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 20/08/2025 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 070793 (form submission date: 18/08/2025); (expected project end date: 25/08/2025).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Luke Whitham
Ethics Admin
Computer Science

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Admin (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Appendix B

Survey Questionnaire

This appendix contains the complete questionnaire used in the user study to gather insights from Thai English learners about their language learning experiences, challenges, and preferences.

B.1 Thai English Learners Questionnaire

The following pages show the Google Forms questionnaire that was distributed to 218 Thai English learners during the requirements gathering phase of this research. The questionnaire was designed to:

- Understand the demographic profile of Thai English learners
- Identify specific challenges faced in English grammar learning
- Assess current tool usage and satisfaction levels
- Gather preferences for educational technology features
- Validate design decisions for the Thaislate tool

The questionnaire consisted of multiple sections covering demographics, learning context, grammar challenges, technology usage, and feature preferences. Responses were collected between July and August 2025 and formed the basis for the requirements analysis presented in Chapter 3.

Thai English Learners Survey – Help Us Improve Language Tools

Research Project Title and Researchers
ชื่อโครงการวิจัยและนักวิจัย

Project Title: Context-Aware Thai-English Translation with Integrated Grammar Learning: An LLM-Based Educational Tool

ชื่อโครงการ: การแปลภาษาไทย-อังกฤษที่เข้าใจบริบทพร้อมการเรียนรู้ไวยากรณ์แบบบูรณาการ: เครื่องมือการศึกษาที่ใช้โมเดลภาษาขนาดใหญ่

Lead Researcher: Papatchol Thientong (นาย ปภัสชล เทียนทอง)
Student Researcher, MSc Computer Science (นักศึกษาวิจัย ปริญญาโท วิทยาการคอมพิวเตอร์)
Email: pthientong1@sheffield.ac.uk

Supervisor: Ms. Varvara Papazoglou (อาจารย์ที่ปรึกษา)
Email: v.papazoglou@sheffield.ac.uk
Department of Computer Science, University of Sheffield (ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชฟฟิลด์)

* Indicates required question

Project Information (ข้อมูลโครงการ) และ Research Consent (การยินยอมเข้าร่วมงานวิจัย)

ข้อมูลสรุปโครงการวิจัย
ชื่อโครงการ: การแปลภาษาไทย-อังกฤษที่เข้าใจบริบทพร้อมการเรียนรู้ไวยากรณ์แบบบูรณาการ: เครื่องมือการศึกษาที่ใช้โมเดลภาษาขนาดใหญ่

นักวิจัย: ปภัสชล เทียนทอง (นักศึกษาปริญญาโท), Ms. Varvara Papazoglou (อาจารย์ที่ปรึกษา)
สถาบัน: ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชฟฟิลด์
การศึกษานี้เกี่ยวกับอะไร?

เรากำลังพัฒนาเครื่องมือเว็บไซต์เพื่อการศึกษาที่ช่วยผู้เรียนภาษาอังกฤษชาวไทยปรับปรุงทักษะไวยากรณ์ เครื่องมือนี้แปลประโยคภาษาไทยเป็นภาษาอังกฤษและให้คำอธิบายเกี่ยวกับการเลือกใช้ไวยากรณ์และรูปแบบต่าง ๆ เราต้องการเข้าใจว่าเทคโนโลยีดังกล่าวมีประโยชน์สำหรับผู้เรียนภาษาอังกฤษชาวไทยแค่ไหน
ท่านต้องทำอะไร?

การเข้าร่วมของท่านอาจเกี่ยวข้องกับหนึ่งหรือทั้งสองระยะ:

- **ระยะที่ 1:** ตอบแบบสอบถามออนไลน์เกี่ยวกับประสบการณ์การเรียนภาษาอังกฤษ (5-10 นาที)
- **ระยะที่ 2:** ทดสอบเครื่องมือแปลภาษาและให้ความคิดเห็น (30 นาที)

เวลารวม: 5-10 นาที (เฉพาะระยะที่ 1) หรือ 40 นาที (ทั้งสองระยะ)
ข้อมูลสำคัญ

- การเข้าร่วมเป็นการสมัครใจโดยสมบูรณ์
- ท่านสามารถถอนตัวได้ตลอดเวลาโดยไม่ต้องให้เหตุผล
- เฉพาะอายุและรหัสผู้เข้าร่วมที่ท่านสร้างเองเท่านั้นที่จะถูกเก็บเป็นข้อมูลส่วนบุคคล
- คำตอบทั้งหมดจะถูกเก็บเป็นความลับและไม่ระบุตัวตน
- ข้อมูลจะถูกเก็บอย่างปลอดภัยและใช้เพื่อวัตถุประสงค์การวิจัยเท่านั้น

สิทธิของท่าน

ท่านมีสิทธิในการเข้าถึง แก้ไข หรือขอให้ลบข้อมูลส่วนบุคคลของท่าน สำหรับข้อมูลการปกป้องข้อมูลแบบเต็ม กรุณาดาวน์โหลดแบบฟอร์มข้อมูลฉบับสมบูรณ์ [Project Information] [Consent Form]

ข้อมูลติดต่อ

นักวิจัยหลัก: ปภัสชล เทียนทอง - pthientong1@sheffield.ac.uk

อาจารย์ที่ปรึกษา: Ms. Varvara Papazoglou - v.papazoglou@sheffield.ac.uk

ติดต่อด้านจริยธรรม: Prof. Jon Barker - j.p.barker@sheffield.ac.uk

1. **Taking Part in the Project** การเข้าร่วมโครงการ *

1. ข้าพเจ้าได้อ่านและเข้าใจเอกสารข้อมูลโครงการลงวันที่ 3 กรกฎาคม พ.ศ. 2568 หรือได้รับการอธิบายโครงการอย่างครบถ้วนแล้ว (หากคุณตอบ "ไม่ใช่" ต่อคำถามนี้ โปรดอย่ากรอกส่วนที่เหลือของแบบฟอร์มจนกว่าคุณจะเข้าใจว่าการเข้าร่วมโครงการหมายความว่าอย่างไร)

2. ข้าพเจ้าได้รับโอกาสในการถามคำถามเกี่ยวกับโครงการ

3. ข้าพเจ้าเข้าใจวิธีการแสดงความกังวลหรือการร้องเรียน

4. ข้าพเจ้าตกลงเข้าร่วมระยะที่ 1: การสำรวจการเรียนภาษาอังกฤษทั่วไป (ประมาณ 5-10 นาที)

5. ข้าพเจ้าเข้าใจว่าสามารถเลือกเข้าร่วมหนึ่งหรือทั้งสองระยะของการวิจัยได้

6. ข้าพเจ้าเข้าใจว่าจะสร้างรหัสผู้เข้าร่วมของตนเองเพื่อปกป้องการไม่ระบุตัวตนในขณะที่ยอมให้เข้าร่วมระยะที่ 2

7. ข้าพเจ้าเข้าใจว่ามีเพียงอายุและรหัสผู้เข้าร่วมที่สร้างขึ้นเองเท่านั้นที่จะถูกเก็บรวบรวมเป็นข้อมูลส่วนบุคคลสำหรับการศึกษานี้

8. ข้าพเจ้าเข้าใจว่าการเลือกที่จะเข้าร่วมเป็นอาสาสมัครในการวิจัยนี้ไม่ได้สร้างข้อตกลงที่มีผลผูกพันทางกฎหมายและไม่มีเจตนาที่จะสร้างความสัมพันธ์ในการจ้างงานกับมหาวิทยาลัยเชฟฟิลด์

9. ข้าพเจ้าเข้าใจว่าสามารถถอนตัวจากการวิจัย/การศึกษาได้ตลอดเวลา ข้าพเจ้าเข้าใจว่าไม่จำเป็นต้องให้เหตุผลใด ๆ และจะไม่มีผลเสียหายใด ๆ หากเลือกที่จะถอนตัว

10. ข้าพเจ้าเข้าใจว่าหากเลือกที่จะถอนตัว อาจไม่สามารถลบข้อมูลการวิจัยที่ข้าพเจ้าได้มีส่วนร่วมได้หากถอนตัวหลังจากวันที่ 1 สิงหาคม พ.ศ. 2568

Mark only one oval.

Yes (ใช่)

No (ไม่ใช่)

2. **How my information will be used during and after the project** *
วิธีการใช้ข้อมูลของข้าพเจ้าในระหว่างและหลังโครงการ

1. ข้าพเจ้าเข้าใจว่ารายละเอียดส่วนบุคคลของข้าพเจ้า เช่น ชื่อ หมายเลขโทรศัพท์ ที่อยู่ และที่อยู่อีเมล ฯลฯ จะไม่ถูกเห็นโดยบุคคลใดนอกเหนือจากทีมโครงการ

2. ข้าพเจ้าเข้าใจว่าคำตอบของข้าพเจ้าต่อแบบสอบถามทั้งสองชุดอาจถูกอ้างอิงแบบไม่เปิดเผยชื่อในสิ่งพิมพ์ รายงาน หน้าเว็บ และผลงานวิจัยอื่นๆ ข้าพเจ้าเข้าใจว่าชื่อของข้าพเจ้าจะไม่ถูกระบุในผลงานเหล่านี้

3. ข้าพเจ้าเข้าใจว่าข้อมูลแบบสอบถามที่ไม่ระบุชื่ออาจถูกใช้สำหรับการวิเคราะห์การวิจัยและการตีพิมพ์

4. ข้าพเจ้าเข้าใจว่าจะไม่มีการบันทึกวิดีโอระหว่างการศึกษานี้

Mark only one oval.

Yes (ใช่)
No (ไม่ใช่)

Basic Information / ข้อมูลพื้นฐาน

3. Age Group / ช่วงอายุของคุณ *

Mark only one oval.

- 18–25
- 26–35
- 36–45
- 46+

4. Current Status / สถานภาพปัจจุบัน *

Mark only one oval.

- University student / นักศึกษา
- English teacher or tutor / ครู หรือ ติวเตอร์ภาษาอังกฤษ
- Other teacher/educator / ครู/นักการศึกษาสาขาอื่น
- Office worker/Corporate employee / พนักงานออฟฟิศ/บริษัท
- Government employee / ข้าราชการ
- Healthcare professional / บุคลากรทางการแพทย์
- Business owner/Entrepreneur / เจ้าของธุรกิจ/ผู้ประกอบการ
- Freelancer/Self-employed / ฟรีแลนซ์/ประกอบอาชีพอิสระ
- Job seeker/Between jobs / กำลังหางาน/ว่างงาน
- Homemaker / แม่บ้าน/พ่อบ้าน
- Retired / เกษียณแล้ว
- Other: _____

5. How long have you been using English? / คุณใช้ภาษาอังกฤษมานานเท่าไหร่แล้ว? (ตั้งแต่เริ่มเรียนภาษา หรือ ใช้ในชีวิตประจำวัน) *

Mark only one oval.

- Rarely used English / ใช้ภาษาอังกฤษน้อยมาก
- Less than 5 years / น้อยกว่า 5 ปี
- 5-10 years / 5-10 ปี
- 11-15 years / 11-15 ปี
- 16-20 years / 16-20 ปี
- More than 20 years / มากกว่า 20 ปี

6. How would you rate your current English level? / คุณจะประเมินระดับภาษาอังกฤษปัจจุบันของคุณอย่างไร? *

Mark only one oval.

- Beginner / เริ่มต้น (A1) - ใช้ประโยคพื้นฐานเกี่ยวกับตัวเอง สามารถแนะนำตัวและถามตอบข้อมูลส่วนตัวง่ายๆ
- Elementary / ระดับพื้นฐาน (A2) - เข้าใจประโยคง่ายๆ เกี่ยวกับครอบครัว การซื้อของ การทำงาน สามารถสื่อสารในสถานการณ์ทั่วไป
- Intermediate / ปานกลาง (B1) - เข้าใจประเด็นหลักในเรื่องที่คุ้นเคย สามารถเขียนข้อความง่ายๆ เกี่ยวกับหัวข้อที่สนใจ
- Upper-Intermediate / ปานกลางชั้นสูง (B2) - เข้าใจข้อความที่ซับซ้อน สามารถสนทนาได้อย่างคล่องตัว เขียนข้อความที่มีรายละเอียด
- Advanced / สูง (C1) - เข้าใจข้อความยาวและซับซ้อน ใช้ภาษาได้อย่างยืดหยุ่นในสถานการณ์ทางสังคม วิชาการ และการทำงาน
- Near-native / เกือบเจ้าของภาษา (C2) - เข้าใจทุกสิ่งที่อ่านและฟังได้อย่างง่ายดาย สามารถแสดงความคิดเห็นที่ซับซ้อนได้อย่างแม่นยำ

7. Comments regarding any questions in this section / ข้อความชี้แจงเพิ่มเติมเกี่ยวกับคำถามแต่ละข้อในส่วนนี้

_____

_____

_____

_____

_____

Current Learning Methods / วิธีการเรียนปัจจุบัน

8. How do you currently engage with English? (Select all that apply) * / คุณมีการใช้หรือเกี่ยวข้องกับภาษาอังกฤษในปัจจุบันอย่างไร? (เลือกได้หลายข้อ)

Tick all that apply.

- [ ] Formal classes (university/college/institute) / เรียนในสถาบัน (มหาวิทยาลัย/วิทยาลัย/สถาบันภาษา)
- [ ] Private tutoring / เรียนพิเศษ
- [ ] Online courses/apps / คอร์สออนไลน์/แอปฯ
- [ ] Self-study with textbooks / เรียนด้วยตนเองจากตำรา
- [ ] Watching English movies/series / ดูหนัง/ซีรีส์ภาษาอังกฤษ
- [ ] Reading English books/articles/websites / อ่านหนังสือ/บทความ/เว็บไซต์ภาษาอังกฤษ
- [ ] Listening to English music/podcasts / ฟังเพลง/พอดแคสต์ภาษาอังกฤษ
- [ ] Using English at work / ใช้ภาษาอังกฤษในการทำงาน
- [ ] Speaking with native speakers/foreigners / พูดกับเจ้าของภาษา/ชาวต่างชาติ
- [ ] Playing English games/social media / เล่นเกม/ใช้โซเชียลมีเดียภาษาอังกฤษ
- [ ] Professional development/training / พัฒนาทักษะเพื่อการทำงาน
- [ ] Rarely use English / ใช้ภาษาอังกฤษน้อยมาก
- [ ] No longer use English / ไม่ได้ใช้ภาษาอังกฤษแล้ว
- [ ] Other: _____

9. What is/was your primary motivation for learning English? / แรงจูงใจหลักในการเรียนภาษาอังกฤษของคุณคืออะไร (เมื่อตอนที่ยังเรียน)? *

Tick all that apply.

- [ ] Career advancement / ความก้าวหน้าในการทำงาน
- [ ] Academic purposes / วัตถุประสงค์ทางวิชาการ
- [ ] Travel and tourism / การเดินทางและท่องเที่ยว
- [ ] Personal interest / ความสนใจส่วนตัว
- [ ] International communication / การสื่อสารระหว่างประเทศ
- [ ] Immigration requirements / ข้อกำหนดการอพยพ
- [ ] Exams and/or school requirements / ข้อสอบ และ/หรือ โรงเรียน

10. How effective are the ways you currently engage with English for improving your English skills? / วิธีที่คุณใช้ภาษาอังกฤษในปัจจุบันช่วยพัฒนาทักษะของคุณได้ดีแค่ไหน? *

Mark only one oval.

- ◯ 1 - Not Effective at all / ไม่ช่วยพัฒนาเลย
- ◯ 2 - Slightly effective / ช่วยพัฒนาเล็กน้อย
- ◯ 3 - Moderately effective / ช่วยพัฒนาปานกลาง
- ◯ 4 - Very effective / ช่วยพัฒนามาก
- ◯ 5 - Extremely effective / ช่วยพัฒนามากที่สุด
- ◯ N/A - Rarely use English / ใช้ภาษาอังกฤษน้อยมาก

11. Comments regarding any questions in this section / ข้อความชี้แจงเพิ่มเติมเกี่ยวกับคำถามแต่ละข้อในส่วนนี้

English Learning Challenges / ความท้าทายในการเรียนภาษาอังกฤษ

12. Which aspects of English do you find most challenging? (Select all that apply) / ด้านใดของภาษาอังกฤษที่คุณมองว่าท้าทายที่สุด? (เลือกได้หลายข้อ) *

Tick all that apply.

- [ ] Understanding spoken English / การฟังภาษาอังกฤษ
- [ ] Speaking and pronunciation / การพูดและการออกเสียง
- [ ] Reading comprehension / การอ่านเพื่อความเข้าใจ
- [ ] Writing clearly / การเขียนให้เข้าใจง่าย
- [ ] Vocabulary (not knowing enough words) / คำศัพท์ (รู้คำน้อย)
- [ ] Grammar rules / กฎไวยากรณ์
- [ ] Translating between Thai and English / การแปลระหว่างไทยและอังกฤษ
- [ ] Understanding cultural context / การเข้าใจบริบททางวัฒนธรรม
- [ ] Confidence in using English / ความมั่นใจในการใช้ภาษาอังกฤษ
- [ ] Finding time to practice / การหาเวลาฝึกฝน
- [ ] No major challenges / ไม่มีความท้าทายใหญ่
- [ ] Other: _____

13. How confident do you feel when using English in different situations? / คุณรู้สึกมั่นใจแค่ไหนเมื่อใช้ภาษาอังกฤษในสถานการณ์ต่างๆ? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	5 - Extremely confident / มั่นใจมากที่สุด

14. Describe a specific situation where you had difficulty with English / อธิบายสถานการณ์เฉพาะที่คุณมีปัญหากับภาษาอังกฤษ

For example: writing professional emails, giving presentations, understanding movies, ordering food, job interviews, etc. / ตัวอย่าง: การเขียนอีเมลทางการ การนำเสนอ การดูหนัง การสั่งอาหาร การสัมภาษณ์งาน ฯลฯ

15. Comments regarding any questions in this section / ข้อความชี้แจงเพิ่มเติมเกี่ยวกับคำถามแต่ละข้อในส่วนนี้

Technology and Learning Tools / เทคโนโลยีและเครื่องมือการเรียน

16. How comfortable are you with using technology for learning? / คุณรู้สึกสะดวกสบายแค่ไหนกับการใช้เทคโนโลยีเพื่อการเรียนรู้? *

Mark only one oval.

- ◯ Very uncomfortable / ไม่สะดวกสบายเลย
- ◯ Somewhat uncomfortable / ไม่ค่อยสะดวกสบาย
- ◯ Neutral / เฉยๆ
- ◯ Comfortable / สะดวกสบาย
- ◯ Very comfortable / สะดวกสบายมาก

17. How often do you use web-based learning tools? / คุณใช้เครื่องมือการเรียนรู้บนเว็บบ่อยแค่ไหน? *

Mark only one oval.

- ◯ Daily / ทุกวัน
- ◯ Several times a week / หลายครั้งต่อสัปดาห์
- ◯ Once a week / สัปดาห์ละครั้ง
- ◯ Once a month / เดือนละครั้ง
- ◯ Rarely / นานๆ ครั้ง
- ◯ Never / ไม่เคย

18. Which devices do you use for English learning? / คุณใช้อุปกรณ์ใดในการเรียนภาษาอังกฤษ? *

Tick all that apply.

- [ ] Smartphone / สมาร์ทโฟน
- [ ] Tablet / แท็บเล็ต
- [ ] Laptop / แล็ปท็อป
- [ ] Desktop computer / คอมพิวเตอร์ตั้งโต๊ะ
- [ ] None / ไม่ใช้

19. Have you used any of these translation tools before? / คุณเคยใช้เครื่องมือแปลภาษาเหล่านี้มาก่อนหรือไม่? *

Tick all that apply.

- [ ] Google Translate
- [ ] Microsoft Translator
- [ ] LINE Dictionary
- [ ] Longdo Dictionary
- [ ] ChatGPT or other AI language tools / ChatGPT หรือเครื่องมือภาษา AI อื่นๆ
- [ ] Other translation apps / แอปแปลภาษาอื่นๆ
- [ ] Never used translation tools / ไม่เคยใช้เครื่องมือแปลภาษา

20. How helpful do you find translation tools for learning English? / คุณคิดว่าเครื่องมือแปลภาษามีประโยชน์แค่ไหนในการเรียนภาษาอังกฤษ? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Very helpful / มีประโยชน์มาก

21. How confident are you in trying new online learning tools? / คุณมั่นใจแค่ไหนในการลองใช้เครื่องมือการเรียนออนไลน์ใหม่ๆ? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Very confident / มั่นใจมาก

22. Have you used any AI-powered tools before? / คุณเคยใช้เครื่องมือที่ขับเคลื่อนด้วย AI มาก่อนหรือไม่? *

Mark only one oval.

- ◯ Yes / เคย
- ◯ No / ไม่เคย

23. Comments regarding any questions in this section / ข้อความชี้แจงเพิ่มเติมเกี่ยวกับคำถามแต่ละข้อในส่วนนี้

Learning Preferences / ความชอบในการเรียน

24. When you make a grammar mistake, what kind of feedback do you prefer? / เมื่อคุณทำผิดไวยากรณ์ คุณชอบการแก้ไขแบบไหน? *

Tick all that apply.

- [ ] Just show the correct answer / แค่แสดงคำตอบที่ถูกต้อง
- [ ] Explain why it's wrong / อธิบายว่าทำไมถึงผิด
- [ ] Show examples of correct usage / แสดงตัวอย่างการใช้ที่ถูกต้อง
- [ ] Provide practice exercises / ให้แบบฝึกหัดเพิ่มเติม

25. How do you prefer to learn grammar rules? / คุณชอบเรียนกฎไวยากรณ์อย่างไร? *

Tick all that apply.

- [ ] Through examples and practice / ผ่านตัวอย่างและการฝึกฝน
- [ ] Through explanation of rules / ผ่านการอธิบายกฎ
- [ ] Through visual aids (charts, diagrams) / ผ่านสื่อภาพ (แผนภูมิ แผนผัง)
- [ ] Through interactive exercises / ผ่านแบบฝึกหัดแบบโต้ตอบ
- [ ] Through comparison with Thai language / ผ่านการเปรียบเทียบกับภาษาไทย
- [ ] Through real-world context and applications / ผ่านบริบทและการประยุกต์ใช้จริง

26. How important is it to understand WHY a grammar rule works, not just HOW to use it? / การเข้าใจว่าทำไมกฎไวยากรณ์ถึงใช้แบบนั้น (ไม่ใช่แค่รู้วิธีใช้) สำคัญแค่ไหน? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Very important / สำคัญมาก

27. How much time can you typically dedicate to English learning per week? / โดยทั่วไปคุณสามารถอุทิศเวลาเรียนภาษาอังกฤษได้กี่ชั่วโมงต่อสัปดาห์? *

Mark only one oval.

- ◯ Less than 2 hours / น้อยกว่า 2 ชั่วโมง
- ◯ 2-5 hours / 2-5 ชั่วโมง
- ◯ 6-10 hours / 6-10 ชั่วโมง
- ◯ 11-15 hours / 11-15 ชั่วโมง
- ◯ More than 15 hours / มากกว่า 15 ชั่วโมง

28. Comments regarding any questions in this section / ข้อความชี้แจงเพิ่มเติมเกี่ยวกับคำถามแต่ละข้อในส่วนนี้

Future Tool Expectations / ความคาดหวังเกี่ยวกับเครื่องมือในอนาคต

29. If there was a tool that could translate Thai sentences to English AND explain the grammar choices, what features would you want it to have? / หากมีเครื่องมือที่สามารถแปลประโยคไทยเป็นอังกฤษ และอธิบายตัวเลือกทางไวยากรณ์ได้ คุณต้องการให้มีคุณสมบัติอะไรบ้าง?

30. Which of these features would be most useful to you? / คุณสมบัติใดต่อไปนี้จะมีประโยชน์กับคุณมากที่สุด? *

Tick all that apply.

- [ ] Step-by-step grammar explanations / คำอธิบายไวยากรณ์แบบทีละขั้นตอน
- [ ] Alternative translation options / ตัวเลือกการแปลที่หลากหลาย
- [ ] Comparison with Thai sentence structure / การเปรียบเทียบกับโครงสร้างประโยคไทย
- [ ] Interactive practice exercises / แบบฝึกหัดแบบโต้ตอบ
- [ ] Audio pronunciation / การออกเสียง
- [ ] Progress tracking / การติดตามความก้าวหน้า
- [ ] Difficulty levels / ระดับความยาก
- [ ] Mobile-friendly design / ออกแบบให้เหมาะกับมือถือ
- [ ] Professional/Academic writing focus / เน้นการเขียนเชิงอาชีพ/วิชาการ
- [ ] Context-specific examples / ตัวอย่างเฉพาะบริบท

31. How likely would you be to use such a tool regularly for your English learning? / คุณมีแนวโน้มที่จะใช้เครื่องมือเช่นนี้เป็นประจำในการเรียนภาษาอังกฤษแค่ไหน? *

Mark only one oval.

	1	2	3	4	5	
Very	◯	◯	◯	◯	◯	Very likely / มีแนวโน้มมาก

32. Comments regarding any questions in this section / ข้อความชี้แจงเพิ่มเติมเกี่ยวกับคำถามแต่ละข้อในส่วนนี้

Additional Comments / ความเห็นเพิ่มเติม

33. Any additional comments about your English learning experience or suggestions for educational tools? / ความเห็นเพิ่มเติมเกี่ยวกับประสบการณ์การเรียนภาษาอังกฤษของคุณ หรือข้อเสนอแนะสำหรับเครื่องมือการศึกษา?

Phase 2 Participation Interest ความสนใจในการเข้าร่วมระยะที่ 2

Important: This section is about Phase 2 (Tool Testing and Evaluation) which will take place in August 2025.

สำคัญ: ส่วนนี้เกี่ยวกับระยะที่ 2 (การทดสอบและประเมินเครื่องมือ) ซึ่งจะดำเนินการในเดือนสิงหาคม 2568

34. **Are you interested in participating in Phase 2: Tool Testing and Evaluation?** ท่านสนใจเข้าร่วมระยะที่ 2: การทดสอบและประเมินเครื่องมือหรือไม่? *

Phase 2 will involve: ระยะที่ 2 จะประกอบด้วย:
Testing the Thai-English translation tool in August (Free testing for 2 days) / การทดสอบเครื่องมือแปลภาษาไทย-อังกฤษ (ทดสอบอิสระในเวลาสองวัน ในช่วงเดือนสิงหาคม)
Completing a feedback questionnaire about the tool / การตอบแบบสอบถามความคิดเห็นเกี่ยวกับเครื่องมือ

Mark only one oval.

- ◯ Yes / ใช่
- ◯ No / ไม่ใช่

Thank You กราบขอบพระคุณ
ขอบคุณที่สละเวลาในการตอบแบบสำรวจนี้ คำตอบของท่านมีค่ามากในการพัฒนาเครื่องมือการศึกษาที่ดีขึ้นสำหรับผู้เรียนภาษาอังกฤษชาวไทย

Second Phase Participant Detail / ข้อมูลผู้เข้าร่วมระยะที่ 2

Participant Code Generation การสร้างรหัสผู้เข้าร่วม
To protect your privacy while allowing you to participate in second phase, you will generate your own unique participant code using a simple algorithm based on personal information that cannot be traced back to you:
เพื่อปกป้องความเป็นส่วนตัวของท่าน ในขณะที่ยอมให้ท่านเข้าร่วมระยะที่สอง ท่านจะสร้างรหัสผู้เข้าร่วมเฉพาะตัวของท่านเองโดยใช้อัลกอริธึมง่าย ๆ ที่อิงจากข้อมูลส่วนบุคคลที่ไม่สามารถติดตามกลับมาถึงตัวท่านได้:

1. First letter of your birth month (e.g., January = J, March = M)
   อักษรตัวแรกของเดือนที่เกิด (เช่น มกราคม = J, มีนาคม = M)
2. Second digit of your age (e.g., if 24 years old, use 4)
   หลักที่สองของอายุ (เช่น หากอายุ 24 ปี ใช้ 4)
3. Number of letters in your first name จำนวนตัวอักษรในชื่อแรกในภาษาอังกฤษของท่าน หารด้วย 10 เอาเศษ (เช่นถ้าชื่อภาษาอังกฤษท่านมี 11 ตัว = 1)
4. Number of siblings จำนวนพี่น้อง (ถ้าเกิน 10 เอาหลักหน่วยอย่างเดียว)
5. Last digit of your phone number
   หลักสุดท้ายของหมายเลขโทรศัพท์

Example: If you were born in June, age 23, first name "Somchai" (7 letters), have 2 siblings, and phone number ending in 8, your code would be: J3728
ตัวอย่าง: หากท่านเกิดเดือนมิถุนายน อายุ 23 ชื่อ "Somchai" (7 ตัวอักษร) มีพี่น้อง 2 คน และหมายเลขโทรศัพท์ลงท้ายด้วย 8 รหัสของท่านจะเป็น: J3728

35. **Your Participant Code: (Please enter your calculated 5-character code)** รหัสผู้เข้าร่วมของท่าน: (กรุณากรอกรหัส 5 ตัวอักษรที่คำนวณได้) *

36. **Please provide your preferred contact details:** กรุณาให้รายละเอียดการติดต่อที่ท่านต้องการ: *
Note: This contact information will only be used to coordinate Phase 2 participation and will be deleted immediately after the research is completed.

หมายเหตุ: ข้อมูลการติดต่อนี้จะใช้เฉพาะในการประสานงานการเข้าร่วมระยะที่ 2 และจะถูกลบทันทีหลังจากการวิจัยเสร็จสิ้น

Example: Social media handle / Line ID / mobile phone / Email

เช่น: ชื่อโซเชียล (โปรดระบุ Social Media) / ไอดีไลน์ / เบอร์โทรศัพท์ / อีเมล

รบกวนแคปหน้าจอจำรหัสของตัวเองด้วยนะครับ ขอบพระคุณมาก ๆ ครับ