

Detection and Correction of Homophone Errors in Thai Language using Machine Learning Techniques in Courtroom Transcriptions

Nuttapong Em-im¹ and Sanya Khruahong^{1,*} 

¹Department of Computer Science and Information Technology, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

*Corresponding author: Sanya Khruahong, sanyak@nu.ac.th

Article Information

Article History:

Received: 4 February 2025

Revised: 25 April 2025

Accepted: 24 August 2025

Published: 17 June 2026

Keywords:

Courtroom Transcription

Machine Learning

Natural Language Processing

Speech-to-Text

Thai Homophone Errors

Abstract

Homophone errors in Thai courtroom transcriptions pose significant challenges to ensuring accurate and reliable legal documentation. In legal contexts, such errors can lead to serious misinterpretations that alter the meaning of witness testimonies, legal arguments, or judicial rulings, potentially influencing case outcomes. While prior ASR and NLP research has focused on general Thai speech, few studies have addressed the specific challenges of homophone correction in legal discourse. This study aims to enhance Thai courtroom transcription accuracy by detecting and correcting homophone errors using machine learning (ML) techniques. It integrates Google's Speech-to-Text API with a deep learning-based correction model that utilizes Thai-BERT and Transformer architectures. The dataset comprises 96 hours of annotated courtroom recordings from Thai provincial courts, yielding over 1.2 million tokens and 18,000 labeled homophone instances. Preprocessing involved audio segmentation, noise reduction, and token-level context tagging. The proposed approach combines contextual analysis, a domain-specific Thai Legal Term Dictionary, and post-processing algorithms to refine transcriptions. Experimental results show a 2.63% increase in overall transcription accuracy (from 91.10% to 93.73%) and a homophone correction rate of 71.5%. The model achieves a precision of 89.2%, a recall of 78.3%, and an F1-score of 83.3%. These improvements have significant legal implications, reducing the risk of misinterpretation in judicial proceedings. The findings highlight the potential of integrating ML-based models into courtroom transcription workflows, offering a scalable and automated solution for enhancing legal documentation. Future work will focus on expanding the dataset, improving model adaptability to real-time and multi-speaker environments, and deploying the system in operational legal settings.

1. Introduction

Accurate courtroom transcription is essential for maintaining reliable legal records and ensuring fairness, transparency, and accountability within the judicial system. In Thailand, transcription is still largely performed manually, requiring court staff to listen to and transcribe spoken dialogue. This process is labor-intensive, time-consuming, and susceptible to human error. As legal caseloads increase and legal language becomes more complex, traditional transcription processes face increasing demands. This highlights the need for automated systems capable of handling

the linguistic characteristics of Thai courtroom speech (Phatthiyaphaibun et al., 2022; Suwanbandit et al., 2023).

A particularly important challenge in Thai courtroom transcription is the prevalence of homophones, which are words that share the same pronunciation but differ in meaning. For example, mishearing โจทย์ (/Cothk/ – plaintiff) as โจทย์ (/Cothy/ – problem) alters the meaning of legal statements and may affect case interpretation. Similarly, confusing ข้าพเจ้า (I, myself) with เข้าไป (to go inside) may misrepresent a speaker's intent or role. Such homophone errors, if left uncorrected, may result in inaccurate transcripts and misinterpretation

of judicial information (Burt and Jared, 2016; Trivedi et al., 2018).

These challenges are amplified by the limitations of general-purpose speech-to-text (STT) systems, which are typically trained on non-specialized language corpora. Legal Thai contains archaic terminology, compound expressions, and role-specific language that may be difficult for STT systems to recognize accurately. In addition, homophones present a substantial challenge when contextual information is insufficient for disambiguation (Iancu, 2019; Ali and Renals, 2018). Table 1 presents examples of Thai homophones that are commonly misinterpreted in legal settings.

To address these limitations, this study integrates Google’s Speech-to-Text API with a Transformer-based correction model, incorporating Thai-BERT and a Thai Legal Term Dictionary. The correction process involves contextual analysis, substitution rules, and post-processing to resolve semantic inconsistencies. Machine learning techniques, particularly those based on contextual embeddings, are well suited to this task because they analyze surrounding textual context to infer intended meaning (Hämäläinen et al., 2021).

Transformer-based models such as Thai-BERT were selected over traditional LSTM or CNN models due to their superior contextual understanding. LSTM and CNN approaches, while effective in some low-resource settings, often struggle with tonal complexity and domain-specific terminology. Thai-BERT, pre-trained on large-scale Thai corpora, provides deeper semantic embeddings, enabling more accurate homophone disambiguation. Likewise, Google’s STT API offers strong Thai language support and provides a practical foundation for post-correction enhancements (Iancu, 2019; Hämäläinen et al., 2021).

Beyond the technical contributions, this research directly benefits legal professionals. Judges rely on accurate transcripts for decision-making, lawyers for legal argument construction, and clerks for documenting court proceedings. Automating homophone correction reduces the time and labor associated with manual edits and increases trust in digital legal records. This is especially impactful in high-volume or rural courts where skilled human transcribers may be limited (Ali and Renals, 2018; Kumar et al., 2024).

By developing a domain-aware, ML-based correction framework for Thai courtroom speech, this research aims to reduce transcription errors, improve legal documentation accuracy, and contribute to more transparent and efficient judicial processes.

In response to the challenges posed by homophone errors in Thai legal transcription, this study proposes a domain-specific correction framework combining speech recognition, contextual language modeling, and legal lexicon integration. The following sections present relevant background research, describe the proposed methodology in detail, and evaluate the model’s effectiveness using real-world courtroom data.

2. Related Work

2.1 General Challenges in Thai ASR

Automatic Speech Recognition (ASR) for Thai faces unique linguistic and technical obstacles. Most notably, Thai is a tonal language with five distinct tones, meaning identical phonetic sequences can correspond to entirely different words. This tonal sensitivity significantly complicates phoneme recognition and contributes to frequent homophone confusion, particularly in rapid or noisy speech. Unlike English, where stress primarily affects emphasis, tonal shifts in Thai directly impact word identity and meaning.

Another major challenge is the scarcity of large-scale annotated Thai speech datasets. While high-performing ASR systems typically rely on abundant training data, Thai remains under-resourced, especially in specialized domains such as legal or governmental language. This limitation hinders model generalizability and reduces effectiveness in domain-specific applications such as courtroom transcription.

Structurally, Thai text lacks explicit word boundaries and frequently exhibits syntactic ambiguity. Sentences often omit subjects, pronouns, or markers, and modifiers appear in flexible positions. These characteristics make segmentation, part-of-speech tagging, and contextual inference more complex, yet essential, for accurate transcription and correction (Hämäläinen et al., 2021).

Recent developments such as Thai Wav2Vec2.0, trained on CommonVoice V8, have shown promising improvements over traditional LSTM- and CNN-based architectures in handling Thai speech (Phatthiyaphaibun et al., 2022). In parallel, domain-specific ASR studies in legal contexts, such as courtroom transcription in French, have explored challenges including speaker overlap and acoustic variation, reinforcing the need for robust, context-aware systems (Garneau and Bolduc, 2024).

Given these issues, general-purpose ASR models often fall short in Thai-language tasks that require high precision, particularly in legal transcription. This study addresses these limitations by introducing domain-adapted modeling and contextual correction strategies designed specifically for Thai tonal, structural, and legal characteristics.

2.2 Homophone Errors in Legal Transcription

Homophones pose a significant challenge in Thai legal transcription because of the tonal nature of the language and the presence of many phonetically identical words with different legal meanings. In courtroom settings, such confusion can alter the semantic integrity of key statements and, consequently, affect judicial outcomes. While general ASR models may transcribe phonetically

Table 1. Examples of Thai homophones commonly encountered in legal contexts.

Homophone Pair (Thai)	Incorrect Transcription (Meaning in English)	Correct Term (Meaning in English)	Example in Legal Context
โจทก์ -- โจทย์ /Cothk – cothy/	โจทย์ (problem, question)	โจทก์ (plaintiff)	“The plaintiff (โจทก์) is responsible for providing evidence in the case.”
ข้าพเจ้า -- เข้าไป /Khāphcēā – khēāpi/	เข้าไป (to go inside)	ข้าพเจ้า (I, myself)	“I (ข้าพเจ้า) testify that the contract was signed legally.”
จำเลย -- จำเริญ /Cāley – cāreiy/	จำเริญ (to flourish, to prosper)	จำเลย (defendant)	“The defendant (จำเลย) denies all charges brought against them.”
พยาน -- พยัญชนะ /Phyān – phyaṇṇa/	พยัญชนะ (consonant)	พยาน (witness)	“The court calls the witness (พยาน) to provide testimony.”
คำฟ้อง -- คำร้อง /Khāfxng – khāfxng/	คำร้อง (petition)	คำฟ้อง (lawsuit, complaint)	“The lawsuit (คำฟ้อง) was submitted with supporting evidence.”
สัญญา -- สัญญาณ /Sayyā – sayyān/	สัญญาณ (signal)	สัญญา (contract)	“The parties signed a contract (สัญญา) before proceeding with the agreement.”

correct words, they often fail to select contextually or legally appropriate terms.

For example, the Thai homophone pair โจทก์ (/Cothk/ – plaintiff) and โจทย์ (/Cothy/ – problem) frequently appears in court proceedings. A sentence such as โจทก์ยื่นฟ้อง (the plaintiff filed a lawsuit) could be mistranscribed as โจทย์ยื่นฟ้อง (the problem filed a lawsuit), rendering the statement legally nonsensical. Another example involves the confusion between ข้าพเจ้า (I, a formal legal pronoun) and เข้าไป (to go inside), which may distort the speaker’s role or responsibility in witness testimony.

Such errors are particularly problematic in legal proceedings, where precise wording has implications for liability, witness credibility, and the interpretation of contractual terms. These types of homophone confusions have not been thoroughly addressed by prior ASR systems or correction algorithms, which typically lack domain-specific semantic constraints.

This study directly addresses this challenge by focusing on homophone detection and correction in Thai courtroom transcriptions. By combining contextual modeling with a curated Thai Legal Term Dictionary, the proposed approach aims to reduce semantic errors and ensure that transcription outputs align with legal standards.

2.3 Machine Learning and Contextual Correction Approaches

Recent advancements in Natural Language Processing (NLP) and deep learning have introduced powerful techniques for disambiguating homophones through context-aware modeling. Transformer-based architectures, particularly BERT and its variants, have been

widely adopted because of their ability to capture semantic relationships within surrounding text, making them suitable for tasks such as grammatical error correction, named entity recognition, and contextual prediction (Hämäläinen et al., 2021).

In the context of homophone correction, these models excel at analyzing the relationship between a word and its surrounding context within a sentence or paragraph. For example, BERT-based models have been applied to homophone detection tasks with promising results, particularly when integrated with domain-specific lexicons and fine-tuning strategies. However, most of these applications have been limited to general text processing or well-resourced languages.

For Thai, few studies have applied deep contextualized embeddings to spoken legal discourse. Prior work has focused primarily on sentiment analysis or depression detection in social media (Hämäläinen et al., 2021), which lacks the complexity and domain-specific vocabulary of courtroom speech.

To address this gap, this study implements a correction model based on Thai-BERT, fine-tuned on courtroom transcription data. The model leverages sequence-to-sequence learning to evaluate contextual probability and predict the correct homophone within a given legal sentence. In addition, it is integrated with a Thai Legal Term Dictionary to apply domain-specific substitution rules that align with legal phrasing and usage. This dual-layer correction mechanism, consisting of deep contextual modeling and legal-term validation, provides a robust approach for improving semantic and legal accuracy in Thai ASR outputs.

2.4 Summary and Research Gap

The literature review highlights the unique challenges associated with developing accurate ASR systems for Thai, particularly in legal contexts. Due to tonal complexity, low-resource conditions, and syntactic ambiguity, general-purpose ASR systems often struggle to produce reliable transcriptions. Homophone errors, in particular, remain a persistent challenge when legal terminology is involved. Misinterpretations of key legal terms can significantly affect the meaning of transcripts, influencing case documentation and judicial outcomes.

Previous research has addressed various aspects of ASR and homophone detection in other languages, and some studies have applied Transformer-based models to Thai NLP tasks. However, few studies have explored the combined use of contextual deep learning and domain-specific correction for legal speech. Existing Thai ASR models typically lack integration with legal knowledge bases, making them prone to critical transcription errors in courtroom settings.

This study addresses these gaps by proposing a two-stage correction framework that integrates Thai-BERT for contextual disambiguation and a curated Thai Legal Term Dictionary for domain-aware validation. Unlike prior approaches, the proposed method is tailored to legal discourse and directly targets the homophone-related challenges of Thai courtroom transcription. This integrated strategy aims to improve transcription accuracy while maintaining legal precision, thereby supporting more reliable and scalable ASR solutions for judicial applications.

3. Research Methodology

The overall research methodology is illustrated in Fig. 1. The proposed framework consists of three main stages: speech-to-text conversion using Google's Speech-to-Text API, homophone error detection through contextual analysis, and machine learning-based correction using Thai-BERT and a Thai Legal Term Dictionary. These stages operate sequentially to transform courtroom audio recordings into legally accurate transcriptions while minimizing semantic errors caused by homophone confusion.

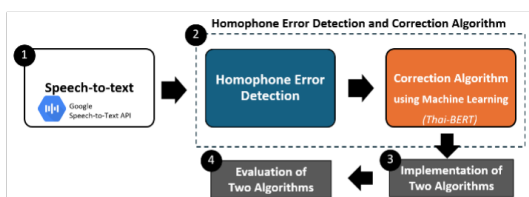


Figure 1. Overview of Research Methodology.

3.1 Speech-to-Text

The first step of the methodology involves converting courtroom speech recordings into textual form using Google's Speech-to-Text API. This API was selected because of its strong performance in speech recognition tasks across multiple languages, including Thai.

1. Courtroom audio recordings are collected and pre-processed to improve audio quality through the application of noise reduction and silence detection techniques.
2. The audio recordings are segmented into smaller chunks to satisfy the processing limitations of Google's API, with each segment not exceeding 60 seconds.
3. Google's Speech-to-Text API processes the segmented audio and generates an initial transcript.
4. The resulting text undergoes basic tokenization and normalization to prepare it for subsequent analysis.

Despite its advantages, Google's Speech-to-Text API lacks domain-specific contextual understanding, which may result in the misinterpretation of Thai homophones and legal terminology. To address this limitation, the next stage of the methodology focuses on homophone error detection and correction.

3.2 Homophone Error Detection and Correction Algorithm

To improve transcription accuracy, this study implements a machine learning-based homophone error detection and correction system consisting of three key techniques.

1. Word Frequency Analysis for Error Detection

A word frequency analysis is performed on the transcribed text to identify unusual or contextually inconsistent words that may indicate homophone errors. This approach involves constructing a frequency distribution of commonly used legal terms in courtroom transcripts and comparing transcribed words with expected word frequencies. By identifying deviations from typical usage patterns, the method helps detect potential misinterpretations and improves the overall accuracy of legal transcriptions.

2. Comparison with a Legal Term Dictionary

A Thai Legal Term Dictionary is incorporated to validate and correct homophones commonly encountered in legal contexts. The dictionary contains a comprehensive collection of legal terms and their frequently misinterpreted counterparts derived from historical legal transcripts. In addition, it includes context-aware substitution rules

that identify and correct homophone errors based on their positions within legal sentence structures, thereby improving transcription accuracy. Domain-adapted models such as LegalBERT-th demonstrate the benefits of combining Transformer architectures with legal-specific vocabulary and document structures (Wiratchawa et al., 2021).

3. Machine Learning-Based Prediction Using Transformer Models

To improve homophone correction accuracy, this study fine-tunes a Transformer-based model, Thai-BERT, to identify the most contextually appropriate word replacements (Hämäläinen et al., 2021). By leveraging contextual embeddings, the model analyzes surrounding words to infer the intended meaning of each homophone. It also applies sequence-to-sequence learning, enabling consistent corrections across multi-sentence contexts, which is particularly important for maintaining coherence in courtroom transcripts. Recent developments such as PhayaThaiBERT, which extends Thai-BERT to support unassimilated loanwords, have further improved the ability of Thai language models to interpret complex linguistic patterns found in legal documents (Sriwirote et al., 2023).

Thai-BERT was selected over multilingual or general-purpose Transformer models because of its specialized pretraining on large-scale Thai corpora that include both colloquial and formal language. This characteristic makes it well suited for capturing the syntactic and semantic characteristics of Thai, particularly when interpreting legal expressions that require precise contextual understanding. Compared with multilingual models such as mBERT and XLM-R, Thai-BERT provides richer contextual representations for Thai-specific language structures, which are essential for distinguishing homophones in complex legal sentences.

3.3 Implementation of Two Algorithms

This study develops and compares two algorithms to evaluate the effectiveness of machine learning-based homophone correction.

1. Algorithm 1: Standard Speech-to-Text Transcription (Baseline)

This algorithm utilizes Google's Speech-to-Text API without applying any additional correction mechanisms. It generates raw transcriptions that may contain homophone misinterpretations and other recognition errors. The resulting transcripts serve as a baseline for comparison with the proposed machine learning-enhanced approach, enabling an assessment of the effectiveness of the homophone detection and correction techniques.

2. Algorithm 2: Machine Learning-Based Homophone Correction

This algorithm improves transcription accuracy by integrating machine learning techniques with Google's Speech-to-Text API. After the initial transcription is generated, word frequency analysis is performed to identify potential homophone errors. The transcribed words are then compared with entries in a Thai Legal Term Dictionary to detect inconsistencies. To further improve accuracy, a Transformer-based model is employed to predict and correct homophone errors using contextual information. This process produces more accurate and reliable courtroom transcriptions by reducing semantic errors associated with homophone confusion.

3.4 Evaluation of Two Algorithms

The performance of both algorithms is evaluated using multiple metrics to assess transcription accuracy and the effectiveness of homophone correction.

1. Word Error Rate (WER)

WER measures the accuracy of transcription by comparing the generated text with manually verified transcripts (Ali and Renals, 2018). It is calculated using the following equation:

$$WER = \frac{S + D + I}{N} \quad (1)$$

where

- S = Number of substitutions (incorrectly transcribed words)
- D = Number of deletions (omitted words)
- I = Number of insertions (additional incorrectly recognized words)
- N = Total number of words in the reference transcript

A perfect transcription yields $WER = 0$, indicating that no transcription errors are present. Higher WER values indicate lower transcription accuracy. In this study, WER is used to compare Algorithm 1 (Baseline STT) and Algorithm 2 (ML-Based Correction) to quantify improvements in transcription accuracy after homophone correction.

2. Homophone Correction Accuracy (HCA)

HCA evaluates the effectiveness of homophone detection and correction (Burt and Jared, 2016). It is calculated using the following equation:

$$HCA = \frac{C}{T} \times 100 \quad (2)$$

where

- C = Number of corrected homophone errors
- T = Total number of homophone errors detected in the initial transcription

A higher HCA value indicates better performance in identifying and correcting homophone errors. If all detected homophone errors are corrected successfully, the system achieves 100% homophone correction accuracy.

3. Precision, Recall, and F1-Score

Precision, Recall, and F1-Score are used to evaluate the ability of the machine learning model to identify and correct homophone errors (Agrawal et al., 2024; Kumar et al., 2024).

Precision is calculated as

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

where

- TP (True Positives) = Number of homophone errors correctly detected and corrected
- FP (False Positives) = Number of incorrect corrections

A higher Precision value indicates that the model performs fewer incorrect corrections.

Recall is calculated as

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

where

- FN (False Negatives) = Number of homophone errors that remain undetected and uncorrected

A higher Recall value indicates that the model successfully detects and corrects more homophone errors.

The F1-Score is calculated as

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

The F1-Score represents the harmonic mean of Precision and Recall, providing a unified measure of overall model performance.

In this study, these metrics are used to compare the performance of Algorithm 1 (Baseline STT) and Algorithm 2 (ML-Based Correction) in evaluating improvements in Thai courtroom transcription accuracy. Although Google's Speech-to-Text API provides robust and scalable speech recognition capabilities, its use as a baseline introduces potential bias. The model is trained

on general-domain data and is not specifically optimized for Thai legal terminology, which may result in lower performance in legal contexts. Furthermore, the API may reflect pronunciation patterns and dialectal variations derived from broad training corpora that do not necessarily correspond to courtroom speech. These limitations may lead to an overestimation of the improvements achieved by the proposed correction model because baseline errors are not uniformly distributed across speech types and contexts. To mitigate this issue, the evaluation focuses specifically on homophone-related corrections to provide meaningful comparative insights.

3.5 Dataset Description and Model Training Details

The dataset used in this study was compiled from real-world courtroom audio recordings collected from the Provincial Court of Tak, Thailand, under a data-sharing agreement. The corpus consists of 96 hours of annotated Thai courtroom speech recorded between 2018 and 2022. All personally identifiable information was anonymized prior to use.

The dataset was manually transcribed and validated by professional legal clerks to ensure transcription reliability. Each transcript includes annotations of homophone occurrences and their corresponding semantic interpretations. After preprocessing, including silence removal, normalization, and tokenization, the dataset contained approximately 1.2 million labeled Thai words, with approximately 18,000 homophone instances identified across 426 unique homophone pairs.

For model training, the corpus was divided into training (70%), validation (15%), and test (15%) sets using stratified sampling to maintain class balance among homophone pairs. Input sequences were segmented using sliding windows of 30 tokens and padded when necessary. Feature extraction was implicitly performed through Thai-BERT embeddings, which capture contextual word representations using Transformer layers.

The Thai-BERT model, based on WangchanBERTa, was fine-tuned using cross-entropy loss and the Adam optimizer. Hyperparameters were optimized through grid search by evaluating learning rates (1×10^{-4} , 2×10^{-5} , and 3×10^{-5}), batch sizes (8, 16, and 32), and dropout rates (0.1 and 0.2). The optimal configuration obtained on the validation set consisted of a learning rate of 2×10^{-5} , a batch size of 16, a dropout rate of 0.1, and six training epochs.

Model performance was evaluated using Precision, Recall, and F1-Score on the test set. In addition, a confusion matrix was employed to analyze error patterns among frequently occurring homophone pairs.

4. Results and Performance Evaluation

This section presents the experimental results and performance evaluation of the two transcription algorithms: Algorithm 1 (Baseline Speech-to-Text) and Algorithm 2 (Machine Learning-Based Homophone Correction). The evaluation focuses on transcription accuracy, homophone detection and correction rates, and the effectiveness of the proposed model in handling Thai homophone errors.

4.1 Comparison of Baseline and ML-Based Transcription Accuracy

The performance of both algorithms was evaluated using Word Error Rate (WER) and Homophone Correction Accuracy (HCA) (Ali and Renals, 2018; Wang et al., 2003). Algorithm 1, which relies solely on Google's Speech-to-Text API, generated raw transcriptions containing a considerable number of homophone misinterpretations. In contrast, Algorithm 2, which integrates machine learning techniques, demonstrated improved performance by correctly identifying and replacing incorrectly recognized homophones in legal transcripts.

Table 2 presents a comparison of the transcription performance achieved by Algorithm 1 and Algorithm 2.

Table 2. Comparison of Algorithm 1 and Algorithm 2 Accuracy.

Algorithm	Accuracy (%)	Homophone Errors	HCA (%)
Algorithm 1 (Baseline)	91.10	471	0.00
Algorithm 2 (ML-Based)	93.73	337	71.50

Table 2 compares the transcription performance of the baseline Google Speech-to-Text system (Algorithm 1) and the proposed machine learning-based correction model (Algorithm 2). The proposed model improved transcription accuracy by 2.63%, while reducing the number of homophone misinterpretations from 471 to 337, representing a reduction of 28.45%.

4.2 Homophone Correction Effectiveness

Table 3 presents the number of homophone errors detected and corrected by the proposed algorithm. The model corrected 337 out of 471 homophone errors, resulting in a correction rate of 71.5%. These corrections are particularly important in legal contexts, where incorrectly transcribed homophones may alter the intended meaning of statements and potentially affect judicial decisions.

Table 3. Percentage of Detected and Corrected Homophone Errors.

Error Type	Occurrences	Detected	Corrected	Rate (%)
Homophone Errors	471	409	337	71.50
Other STT Errors	320	190	128	40.00

The machine learning-based correction model demonstrated strong performance in correcting homophone errors when sufficient contextual information was available. For example, in the phrase

จำเลยให้การว่าโจทก์ไม่มีพยานหลักฐาน

the model successfully corrected โจทก์ (problem) to โจทก์ (plaintiff) based on its contextual relationship with จำเลย and the surrounding legal discourse. Similarly, the model corrected เข้าไป (to go inside) to ข้าพเจ้า (I, myself) in formal testimony contexts, demonstrating accurate semantic interpretation.

However, the model occasionally struggled with ambiguous contexts. For example, in sentences such as

เมื่อคำร้องถูกส่งให้ศาล

the model incorrectly retained คำร้อง (petition) in cases where คำฟ้อง (complaint) was legally more appropriate. This error occurred because of overlapping usage patterns in general administrative and legal contexts. Additional errors were observed for rare or archaic legal terms that were underrepresented in the legal dictionary or pretraining corpus.

These findings suggest that, although the model performs effectively on high-frequency legal expressions and syntactically clear sentences, additional fine-tuning is required to address edge cases involving polysemy, archaic terminology, and syntactic ambiguity.

Overall, the proposed system was particularly effective in handling homophone errors and demonstrated high precision in distinguishing between phonetically similar words within legal transcripts.

4.3 Precision, Recall, and F1-Score Analysis

To evaluate the effectiveness of the proposed model in correcting homophone errors, standard classification metrics were employed, including Precision, Recall, and F1-Score.

The model achieved a Precision of 89.2%, indicating that most corrected homophones corresponded to actual transcription errors. This high Precision value suggests that the model rarely modifies words that are already correct. The Recall value reached 78.3%, demonstrating the model's ability to identify and correct a substantial proportion of homophone errors present in the transcripts. The resulting F1-Score was 83.3%, reflecting a balanced performance between Precision and Recall.

These results demonstrate that the proposed model is effective not only in identifying homophone errors but

also in selecting contextually and legally appropriate corrections. The combination of contextual language modeling and legal-term validation contributed to the strong overall performance observed in the experiments.

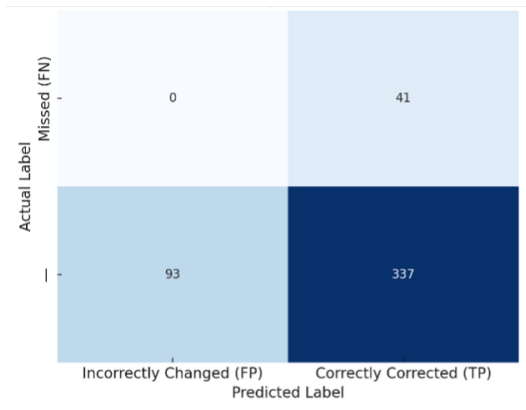


Figure 2. Confusion matrix of the homophone correction model showing 337 correct corrections (TP), 93 missed errors (FN), and 41 incorrect corrections (FP). True negatives are omitted as the evaluation focuses on detected homophone errors.

Fig. 2 presents the confusion matrix of the proposed homophone correction model. Among the detected homophone instances, 337 were correctly corrected (True Positives), 93 errors remained uncorrected (False Negatives), and 41 words were incorrectly modified (False Positives). These results indicate that the model is capable of correcting the majority of homophone errors while maintaining a high level of Precision. Nevertheless, some infrequent homophones and contextually ambiguous cases continue to present challenges.

The absence of True Negatives is attributed to the evaluation design, which focuses exclusively on words identified as potential homophone errors. Consequently, the confusion matrix emphasizes the model's effectiveness in detecting and correcting homophone-related transcription errors rather than its performance on correctly transcribed words.

5. Discussion

The findings of this study highlight the effectiveness of integrating machine learning-based homophone detection and correction into Thai courtroom transcription systems. The experimental results demonstrate that Algorithm 2 (ML-Based Correction) significantly improves transcription accuracy compared with Algorithm 1 (Baseline STT) by reducing homophone misinterpretations and increasing the overall accuracy of legal transcripts.

The integration of word frequency analysis, a legal term dictionary, and a Transformer-based model proved effective in addressing the challenges associated with homophone errors in Thai courtroom settings. The homo-

phone correction rate of 71.5% achieved by Algorithm 2 suggests that contextualized homophone identification, when combined with domain-specific knowledge, can substantially improve the reliability of speech-to-text (STT) systems in legal applications.

Moreover, the 2.63% improvement in transcription accuracy, from 91.10% with the baseline STT system to 93.73% with the ML-based correction model, demonstrates the feasibility of applying post-processing AI techniques to improve ASR-generated transcripts. This improvement is particularly important in legal contexts, where even minor transcription errors may lead to misinterpretations of evidence, testimony, or judicial reasoning.

The findings further confirm that contextual analysis is essential for resolving homophone ambiguity in legal transcription. The Transformer-based model leveraged contextual embeddings to distinguish phonetically similar terms by analyzing surrounding textual information. As a result, the system achieved a Precision of 89.2% and an F1-Score of 83.3%, demonstrating its effectiveness in generating legally appropriate corrections.

The incorporation of a Thai Legal Term Dictionary further enhanced system performance by promoting terminology consistency through domain-specific substitution rules. This hybrid approach ensured that misinterpreted homophones were corrected in accordance with established legal usage.

Beyond technical performance, the proposed system offers practical benefits for legal professionals in courtroom environments. Judges rely on accurate transcripts to support informed decision-making, while attorneys depend on reliable records when constructing legal arguments. Even minor transcription errors, particularly homophone substitutions, may obscure the roles of legal parties or distort legal reasoning. By automating error detection and correction, the proposed system reduces the manual workload of court personnel and supports more efficient documentation processes. Such benefits may be particularly valuable in high-volume or under-resourced courts.

Future improvements may be achieved by incorporating additional linguistic features, including part-of-speech (POS) tagging and syntactic parsing, to improve disambiguation in grammatically complex sentences. Furthermore, integrating speaker identification and discourse-level information may provide richer contextual cues for distinguishing statements made by judges, witnesses, and attorneys.

Although the proposed model demonstrates promising performance in offline settings, deployment in real-time courtroom environments presents additional technical and procedural challenges. Real-time transcription requires low-latency inference, which may necessitate model optimization and hardware acceleration. Furthermore, overlapping speech, background noise, and spontaneous interruptions remain important challenges for maintaining transcription quality. Legal ad-

missibility requirements may also require human oversight and verification. Addressing these issues will require additional engineering efforts, including real-time diarization, speaker segmentation, and confidence-based correction mechanisms to support practical deployment in courtroom environments.

6. Conclusions

This study addressed the challenge of homophone errors in Thai courtroom transcriptions by proposing a machine learning-based correction framework integrated with Google's Speech-to-Text API. The proposed system combines word frequency analysis, a Thai Legal Term Dictionary, and a Transformer-based contextual model (Thai-BERT). Experimental results indicate that the proposed approach improved transcription accuracy from 91.10% to 93.73%, while achieving a homophone correction rate of 71.5%. The model obtained a Precision of 89.2%, a Recall of 78.3%, and an F1-Score of 83.3%, demonstrating its effectiveness in correcting phonetically similar but semantically distinct legal terms.

Despite these promising results, several limitations remain. The model continues to face challenges when handling contextually ambiguous homophones, low-frequency legal expressions, and overlapping speech segments. Addressing these limitations will require expansion of the Thai Legal Speech Corpus, additional legal-domain fine-tuning, and improved speaker diarization techniques for multi-speaker courtroom environments.

Beyond technical contributions, this research supports the development of scalable AI-driven transcription systems for the legal sector. By reducing transcription errors, the proposed framework improves the accuracy of judicial documentation, supports informed decision-making, and reduces the workload of court personnel. These benefits may contribute to improved procedural integrity, legal transparency, and efficient case management, particularly in high-volume or under-resourced courts.

Future work will focus on deploying the proposed framework in real-time courtroom environments and optimizing it for low-latency inference. Additional enhancements may include incorporating role-based semantic information, such as distinguishing speech produced by judges and witnesses, as well as adapting the framework to other tonal or under-resourced languages in Southeast Asia. The proposed approach may also be extended to related legal applications, including attorney-client consultations, police interviews, notarial transcription, and legal education.

By improving the reliability and accessibility of speech-to-text systems in legal environments, this work contributes to the responsible application of artificial intelligence in modern legal infrastructure. This direction is consistent with the growing development of legal NLP resources and benchmarks, including NitiBench for

Thai legal question answering (Akarajaradwong et al., 2025).

CRediT authorship contribution statement

Nuttapong Em-im: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Validation, Visualization, Writing – Original Draft. **Sanya Khruahong:** Conceptualization, Supervision, Methodology, Resources, Project Administration, Validation, Writing – Review & Editing,

Declaration of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Agrawal, P., Sharma, K., Dhage, K., Sharma, I., Rakesh, N., and Kaur, G. (2024). Speech-to-text conversion and text summarization. In *Proceedings of the 2024 First International Conference on Technological Innovations and Advance Computing (TIA-COMP)*, pages 536–541. IEEE. DOI: 10.1109/TIA-COMP64125.2024.00094.
- Akarajaradwong, P., Pothavorn, P., Chaksangchaichot, C., Tasawong, P., Nopparatbundit, T., Pratai, K., and Nutanong, S. (2025). NitiBench: Benchmarking LLM frameworks on Thai legal question answering capabilities. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 34304–34327, Suzhou, China. Association for Computational Linguistics. DOI: 10.18653/v1/2025.emnlp-main.1739.
- Ali, A. and Renals, S. (2018). Word error rate estimation for speech recognition: e-WER. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 20–24, Melbourne, Australia. Association for Computational Linguistics. DOI: 10.18653/v1/P18-2004.
- Burt, J. S. and Jared, D. (2016). The role of lexical expertise in reading homophones. *Quarterly Journal of Experimental Psychology*, 69(7):1302–1321. DOI: 10.1080/17470218.2015.1062528.
- Garneau, N. and Bolduc, O. (2024). The state of commercial automatic French legal speech recognition systems and their impact on court reporters. *arXiv preprint arXiv:2408.11940*.

- Hämäläinen, M., Patpong, P., Alnajjar, K., Partanen, N., and Rueter, J. (2021). Detecting depression in Thai blog posts: A dataset and a baseline. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 20–25, Online. Association for Computational Linguistics. DOI: 10.18653/v1/2021.wnut-1.3.
- Iancu, B. (2019). Evaluating Google speech-to-text API's performance for Romanian e-learning resources. *Informatica Economica*, 23(1):17–25. DOI: 10.12948/issn14531305/23.1.2019.02.
- Kumar, R., Gupta, M., Sharma, P., Soni, N., and Rawat, K. (2024). NLP-based text-to-speech and speech-to-text virtual assistant. In *AIP Conference Proceedings*, volume 3072, page 020031. AIP Publishing. DOI: 10.1063/5.0203299.
- Phatthiyaphaibun, W., Chaksangchaichot, C., Limkonchotiwat, P., Chuangsuwanich, E., and Nutanong, S. (2022). Thai Wav2Vec2.0 with CommonVoice V8 for ASR. *arXiv preprint arXiv:2208.04799*.
- Sriwrote, P., Thapiang, J., Timtong, V., and Rutherford, A. T. (2023). PhayaThaiBERT: Enhancing a pretrained Thai language model with unassimilated loanwords. *arXiv preprint arXiv:2311.12475*.
- Suwanbandit, A., Naowarat, B., Sangpetch, O., and Chuangsuwanich, E. (2023). Thai dialect corpus and transfer-based curriculum learning investigation for dialect automatic speech recognition. In *Proceedings of Interspeech 2023*, pages 2958–2962. DOI: 10.21437/Interspeech.2023-1828.
- Trivedi, A., Pant, N., Shah, P., Sonik, S., and Agrawal, S. (2018). Speech to text and text to speech recognition systems: A review. *IOSR Journal of Computer Engineering*, 20(2):36–43.
- Wang, Y.-Y., Acero, A., and Chelba, C. (2003). Is word error rate a good indicator for spoken language understanding accuracy. In *Proceedings of the 2003 IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 577–582, St. Thomas, U.S. Virgin Islands. IEEE. DOI: 10.1109/ASRU.2003.1318504.
- Wiratchawa, K., Khunthong, T., and Intharah, T. (2021). LegalBERT-th: Development of legal Q&A dataset and automatic question tagging. In *Proceedings of the 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1159–1162, Chiang Mai, Thailand. IEEE. DOI: 10.1109/ECTI-CON51831.2021.9454753.