

Aspect-Level Sentiment Analysis Using WangchanBERTa for Fine-Grained Service Insight Extraction in Hotel Reviews

Thanachok Suwan¹, Manussawee Nokkaew^{1†}, Chayada Surawanitkun¹, Kanda Sorn-In¹,
Nongram Mueanrit¹, Kwankamol Nongpong², Tapanan Yeophantong³,
Wisut Supasai⁴, and Apirat Siritaratiwat⁵, Non-members

ABSTRACT

Online booking site reviews substantially influence Thai SME hotel reputations and consumer decisions. Hotels should readily extract useful information from unstructured Thai-language ratings. WangchanBERTa, a Thai deep learning model, automates hotel sentiment analysis and strategic insight development in this study. System is two-stage. Phase 1 divides 10,040 Thai hotel reviews from Agoda, Booking.com, Traveloka, and Trip.com into good and negative attitudes and determines price, service quality, and cleanliness. Phase 2 extracts aspect-level information across 11 service characteristics to discover complex trends like consumers being satisfied with service but unhappy with cost. The sentiment categorization model performed well with 91.63% accuracy and 89.69% macro F1-score in experiments. The aspect-based sentiment analysis system achieved 91.63% accuracy, 91.55% macro precision, 91.63% recall, 91.52% F1-score, and real-world insight extraction. This methodology helps hoteliers listen to customers, integrate data into business ideas, and compete in Thailand's tourism market.

Keywords: Sentiment Analysis, Aspect-Based Sentiment Analysis, WangchanBERTa, Natural Language Processing, Deep Learning

Manuscript received on December 23, 2025; revised on April 2, 2026; accepted on May 2, 2026. This paper was recommended by Associate Editor Chanon Warisarn.

¹The authors are with Center of Multidisciplinary Innovation Network Talent (MINT Center), Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai Campus, Nong Khai, Thailand.

²The author is with Department of Computer Science Vincent Mary School of Engineering, Science and Technology, Assumption University, Samut Prakan, Thailand.

³The author is with Department of Design and Digital Innovation, Martin de Tours School of Management and Economics, Assumption University, Samut Prakan, Thailand.

⁴The author is with Computer Center Faculty of Science, Ramkhamhaeng University, Bangkok, Thailand.

⁵The author is with Department of IoT and Information Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang, Bangkok, Thailand.

[†]Corresponding author: manuno@kku.ac.th

©2026 Author(s). This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License. To view a copy of this license visit: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Digital Object Identifier: 10.37936/ecti-ec.2026242.262946

1. INTRODUCTION

In the digital economy era, online user reviews have become a critical information source that directly reflects the quality and reputation of hospitality businesses [1], [2], [3]. For small and medium-sized hotel enterprises (SME Hotels), which form the backbone of Thailand's tourism industry, accurately understanding customer feedback is essential for market survival and competitiveness. However, manual analysis of review texts is time-consuming, costly, and susceptible to interpretation bias [3].

Natural Language Processing (NLP) technology offers a promising approach for transforming unstructured textual data into actionable strategic insights [2], [4]. In recent years, Transformer architectures based on self-attention mechanisms have revolutionized NLP by efficiently capturing long-range contextual dependencies. Subsequently, bidirectional encoder models such as Bidirectional Encoder Representations from Transformers (BERT) have been developed and widely adopted as the standard for text analysis tasks [2].

Applying BERT to the Thai language presents unique challenges due to the absence of word boundaries, diverse idiomatic expressions, and context-dependent implicit meanings that models designed primarily for English or other languages cannot adequately capture [1], [5]. To address these limitations, the Vidyasirimedhi Institute of Science and Technology (VISTEC), with support from the Digital Economy Promotion Agency (DEPA), developed WangchanBERTa, a Thai-specific deep learning model pre-trained on over 78 gigabytes of diverse data from websites and social media platforms [1]. This model has been successfully employed in multiple research studies for sentiment analysis of Thai texts, demonstrating significantly higher accuracy than traditional machine learning approaches [4], [5].

In the hotel review domain, prior work reported that traditional methods using Support Vector Machine (SVM) classifiers combined with Delta Term Frequency-Inverse Document Frequency (Delta TF-IDF) weighting achieved approximately 90% accuracy [4]. At the same time, deep learning approaches integrating Thai-specific encoders like WangchanBERTa with Convolutional Neural Network (CNN) / Bidirectional Long Short-Term Memory Network (BiLSTM) architectures yielded significantly

higher Macro F1-scores [6]. However, these methods primarily focus on sentence and document-level classification and lack the granularity required for aspect-level analysis, which is crucial for identifying specific service dimensions such as pricing, cleanliness, room quality, and staff service that influence customer satisfaction.

To bridge this gap, this research develops an automated sentiment analysis and insight extraction system for Thai SME hotels using fine-tuned WangchanBERTa, capable of both sentiment analysis and aspect-based sentiment analysis (ABSA). The system architecture comprises two phases:

Phase 1: Development of a sentiment analysis model by collecting and preparing hotel review data from Agoda, Booking.com, Traveloka, and Trip.com platforms, followed by data cleaning and formatting processes for model training and performance evaluation.

Phase 2: Aspect-level insight extraction by defining critical service dimensions such as cleanliness, pricing, room quality, and staff service, then performing aspect detection and sentiment polarity assignment for these aspects.

The system outputs are synthesized into strategic guidelines that enable hoteliers to formulate targeted service improvement measures with precision and efficiency [9], [10], [11].

2. LITERATURE REVIEW

2.1 Sentiment Analysis

Sentiment analysis refers to the task of identifying the writer's emotional stance in a piece of text, typically categorized as positive, negative, or neutral [4]. Reviews in the hotel domain have shown a clear shift over time in the techniques used for this task, particularly toward deep learning-oriented approaches [12]. Earlier studies often relied on traditional machine learning algorithms such as Naïve Bayes, logistic regression, and support vector machines. Although effective to a degree, these techniques struggle to capture contextual dependencies between words, which limits their performance on complex texts [4], [5].

Language-specific models such as WangchanBERTa, a Thai Transformer-based model [1], offer notable improvements in handling informal and idiomatic Thai expressions.

Namee et al. [6] proposed a hybrid method combining Transformer embeddings with Random Forest classification for aspect-based sentiment analysis. Their approach surpassed sequential neural models such as GRU and RNN in terms of F1-score. Bai et al. [7] found that domain-specific prompting strategies and fine-tuning remain essential.

2.2 Transformer Models and Applications in Thai Language

WangchanBERTa, developed by Lowphansirikul et al. [1] and extensively trained on Thai text, captures contextual

subtleties—such as colloquial expressions and non-standard writing styles—more effectively than general-purpose multilingual BERT variants.

Chantrapornchai and Tunsakul [3] incorporated BERT and SpaCy to extract information across tourism-related content, including hotels and attractions, and demonstrated considerable performance advantages over classical ML approaches. Lapchaicharoenkit [8] extended BERT by integrating Dice Loss to alleviate class imbalance issues, enabling the model to interpret long or fragmented Thai texts more robustly.

2.3 Aspect-Based Sentiment Analysis (ABSA) Research

Aspect-based sentiment analysis aims to break down user opinions into specific dimensions—such as price, service quality, or facilities—allowing a more nuanced interpretation than overall sentiment scores [9]. Hua et al. [9] reviewed ABSA frameworks. Nawawi et al. [10] explored zero-shot approaches for detecting aspects in tourism reviews, showing that models can generalize across domains even without domain-specific training. Kontonatsios et al. [11] introduced a standardized ABSA dataset.

2.4 Related Research in Tourism and Hotel Review Sentiment Analysis

In the Thai context, Suraratchai and Phoomvuthisarn [13] introduced a hybrid WangchanBERTa-CNN-BiLSTM architecture for sentiment classification, demonstrating that contextual embeddings combined with sequential layers can enhance performance over classical machine learning models. Bunyatisai et al. [14] improved sentiment prediction for Thai hotel reviews by integrating emoji-aware emotional cues.

Jensirisak et al. [15] applied BERT-based ABSA to Thai smartphone reviews, illustrating that aspect-based methods are feasible for Thai-language content even outside the hospitality domain. Wanasukapunt and Phimoltare [16] developed a deep-learning framework for classifying Thai social media content. Nokkaew et al. [17] analyzed public sentiment toward large-scale transportation projects using machine learning, deep learning, and ChatGPT-assisted models—reflecting Thailand's broader adoption of modern NLP and LLM techniques in tourism-adjacent analytics.

Wen et al. [18] employed BERT and ERNIE to process hotel-review datasets, demonstrating improved contextual understanding compared with traditional models; however, their work remained centered on polarity classification rather than multi-aspect opinion extraction.

Thomson et al. [19] proposed a hybrid BERT architecture with text augmentation for Indonesian hotel reviews, enhancing robustness to linguistic variability yet limiting their analysis to sentence-level sentiment.

Sahin and Eyupoglu [20] conducted a systematic literature review of ABSA applications within the hospitality

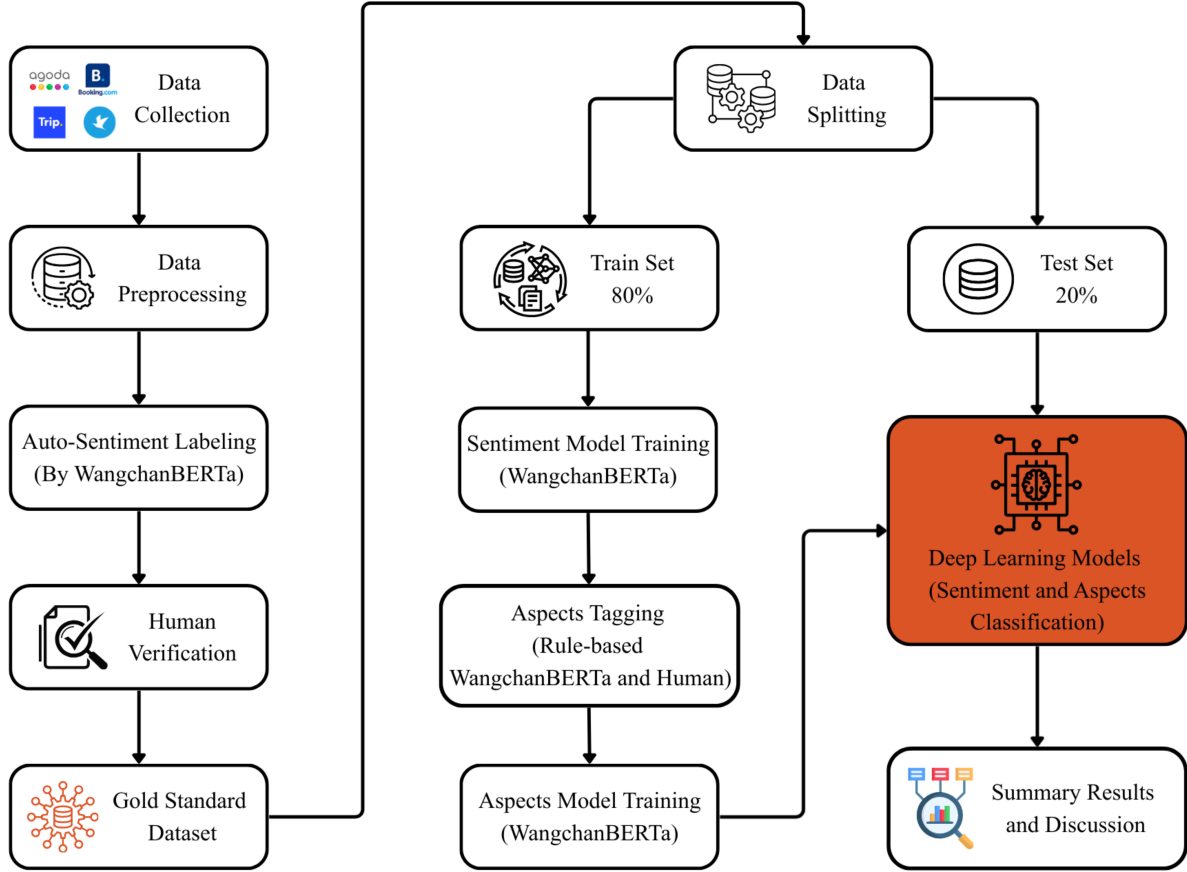


Fig. 1: WangchanBERTa-Based Framework for Aspect-Level Sentiment Analysis of Thai Hotel Reviews.

domain. Their findings reveal persistent challenges—including the scarcity of domain-adapted datasets, inconsistencies in aspect schema design, and the absence of end-to-end ABSA pipelines—thereby reinforcing the limitations observed in works such as Wen et al. [18] and Thomson et al. [19].

Roumeliotis et al. [21] compared BERT-based systems with GPT-Omni models for tourism-related text classification, highlighting the rapidly improving generalization capabilities of LLMs but stopping short of proposing domain-specific pipelines. Nokkaew et al. [22] demonstrated the usefulness of deep learning and ChatGPT-assisted systems for large-scale sentiment studies, though their work focused on transportation projects rather than hotel reviews.

2.5 Research Gaps and Contributions

Despite rapid progress in sentiment analysis, both globally and in Thailand, several limitations remain. Many Thai studies still focus primarily on polarity classification and rarely examine how different service components shape customer experience. Khamphakdee and Seresangtakul [4] and Suraratchai and Phoomvuthisarn [13] overlook multiple service aspects within a single review. Hua et al. [9] highlights the importance of analyzing opinions across multiple dimensions, yet a

number of hotel-review studies (e.g., Wen et al. [18]) continue to prioritize general sentiment rather than aspect-level insights.

Another gap concerns the limited use of WangchanBERTa for Thai hotel ABSA. Its potential for hotel-related aspect classification remains underexplored. Sahin and Eyupoglu [20] notes that the lack of standardized aspect definitions and domain-oriented pipelines remains a challenge.

Finally, most studies lack practical tools for hotel operators.

Contributions of This Work:

- End-to-end ABSA workflow for Thai hotel reviews. A complete pipeline covering data collection from OTAs, preprocessing, automated sentiment labeling and aspect-level modeling, drawing on Hua et al. [9], Nawawi et al. [10], and Thai NLP literature [4], [13].
- Domain-adapted WangchanBERTa model. The model is fine-tuned to classify sentiment across 11 service aspects in Thai hotel reviews, aligning with findings on Transformer effectiveness in hotel domains [18].
- Thai hotel ABSA gold dataset. We construct a validated, domain-specific dataset for reproducible ABSA research.
- Interactive dashboard for hotel operators. Aspect-level predictions are integrated into a dashboard that

provides visual, interpretable insights for Thai SME hotels.

3. METHODOLOGY

3.1 Research Overview

Figure 1 gives an overview of the workflow used in the aspect-based sentiment analysis, which aims to interpret Thai hotel reviews both at the sentiment level and through specific service aspects. The system builds upon modern Transformer-based language models—an approach that has become central in recent NLP developments [1]. WangchanBERTa, in particular, was chosen as the main model because it was pretrained on a very large amount of Thai text and has shown strong performance on Thai-language tasks in earlier studies [8].

The process starts with collecting review messages from several online travel platforms. These reviews are first passed through an automatic module that predicts positive or negative sentiment. Because automated predictions may include errors, the outputs are checked manually, and the corrected version becomes the Gold-Standard Dataset used for training.

After the dataset is finalized, the text is cleaned, normalized, and prepared before being split into training and test sets. From here, the development separates into two paths. One is devoted to training a model for overall sentiment, and the other focuses on aspect identification. Aspect labels are produced through rule-based extraction, assisted by model predictions and refined by human judgment, before being used to train a multi-label classifier.

Both models—the sentiment classifier and the aspect classifier—work together as the core analytical components of the system. Their performance is evaluated using standard quantitative metrics, including Accuracy, Precision, Recall, and F1-Score [22], [23], [24], [25]. In addition, the system incorporates confidence-related measurements to assess prediction reliability and calibration quality [26], [27], [28]. Finally, their results are summarized and discussed, focusing on patterns found in SME hotel reviews and how these insights may support operational decision-making.

3.2 Data Collection

The dataset for this research was collected from four popular online hotel review platforms used in Thailand: Agoda, Booking.com, Traveloka, and Trip.com, which provide authentic user experiences from hotel stays. Data collection was conducted through web scraping of publicly available reviews using keywords related to Thai SME hotels, such as “budget accommodation”, “small hotel”, “good service”, “clean room”, and “delicious breakfast”. Textual reviews, along with supplementary information such as satisfaction ratings, were extracted. All reviews were stored in JSON format and then converted into tabular data for processing convenience. In total, 10,040 review entries were collected. The resulting

consolidated dataset structure is illustrated in Table 1, which presents sample records with key fields including platform, review message, rating, sentiment label, and aspect labels.

3.3 Data Labeling

A semi-automatic labeling approach was employed, combining the WangchanBERTa model with verification by three Thai language experts. Initially, the model automatically classified reviews as positive or negative. Subsequently, each expert independently verified the classifications. Reviews with disagreements among experts or significant deviations from the model’s predictions were excluded, retaining only reviews with 100% consensus among all experts [5].

The final dataset comprises 10,040 entries: 7,071 positive reviews (70.42%) and 2,969 negative reviews (29.58%). Although the dataset exhibits class imbalance, it reflects the realistic user behavior pattern where users tend to provide reviews primarily for positive experiences [10].

3.4 Data Preprocessing

Preprocessing plays a crucial role in improving model quality by reducing textual noise and normalizing variations common in Thai user-generated content. The PyThaiNLP library [29] was selected because it offers reliable Thai language processing utilities, particularly for handling segmentation in a language that lacks explicit word boundaries.

The preprocessing pipeline consists of several stages:

- Text Cleaning
 - Remove punctuation, numbers, symbols, and URLs
 - Eliminate extraneous Unicode characters
 - Correct frequent text anomalies (such as “๕” → “๕”)
- Text Normalization
 - Reduce duplicated characters (“ชอบบบบ” → “ชอบ”)
 - Standardize irregular character forms
- Tokenization

Thai word segmentation is performed using PyThaiNLP’s tokenizer to reflect natural linguistic boundaries.
- Stop-Word Removal

Common function words without sentiment contribution such as “ค่ะ”, “ครับ”, “เลย”, “ก็”, “นะ” are filtered out.

After these steps, the resulting dataset is cleaner, more consistent, and suitable for model training.

3.5 Model architectures

3.5.1 BERT

BERT (Bidirectional Encoder Representations from Transformers) is a language model developed by Devlin et al. at Google AI Language in 2018 [25]. It employs an encoder-only Transformer architecture that learns bidirectional contextual relationships through a self-attention mechanism, enabling a deeper understanding of word usage within sentences and capturing context-

Table 1: Sample data from the multi-source Thai hotel review dataset for ABSA.

Platform	Review Message	Rating	Sentiment Label	Aspect Labels
Agoda	The hotel is very clean, but parking is limited.	7.8	Positive	Cleanliness: Positive, Room: Positive, Atmosphere: Negative
	The accommodation is excellent, with a great location and convenient transportation. The lobby has been newly renovated and looks beautiful. However, the room has a musty smell.	8.5	Positive	Location: Positive, Facilities: Positive, Room: Negative
Booking	Bathroom was new and clean, but the hairdryer didn't work.	7.9	Positive	Cleanliness: Positive, Facilities: Negative
	Breakfast was tasty with good variety, but the dining area was very crowded	7.2	Negative	Food: Positive, Room: Negative
Traveloka	The room is a bit small, but it's reasonably priced.	8.0	Positive	Room: Negative, Price: Positive
	Staff were polite and helpful, but check-in was too slow.	7.0	Negative	Staff: Positive, Check-in: Negative
Trip.com	Great location, but the room is old and has a musty smell.	7.0	Positive	Location: Positive, Room: Negative
	Very weak water pressure and staff were not polite.	6.8	Negative	Facilities: Negative, Staff Service: Negative

dependent meanings more effectively than traditional static embeddings such as word2vec or GloVe.

BERT is pre-trained using two primary tasks: Masked Language Modeling (MLM), in which some tokens in a sentence are masked and the model predicts the missing tokens, and Next Sentence Prediction (NSP), which trains the model to understand inter-sentence relationships for document-level coherence. This pre-training strategy allows BERT to produce flexible contextual embeddings that can be efficiently fine-tuned for downstream tasks such as sentiment analysis, question answering, text classification, and named entity recognition.

BERT achieved state-of-the-art performance on multiple benchmark datasets, including GLUE and SQuAD, and has since become a foundational model in modern NLP. However, subsequent research revealed limitations when applying BERT to languages without explicit word boundaries, such as Thai and Indonesian, motivating adaptations in tokenization strategies and model pre-training for these languages [1]. In Thailand, recent studies have demonstrated increasing adoption of deep contextual models and LLM-integrated frameworks for analyzing social media sentiment and emerging emotional patterns, underscoring the need for models specifically designed to handle the complex linguistic characteristics of Thai [22].

To address these challenges for Thai, WangchanBERTa

was introduced as a Thai-specific Transformer-based language model pre-trained on large-scale Thai corpora, providing stronger coverage of Thai vocabulary and morphology for downstream tasks such as sentiment analysis.

3.5.2 WangchanBERTa

WangchanBERTa is a Thai-specific deep language model developed by AI Research Thailand in 2021 [1]. It is based on the Robustly Optimized BERT (RoBERTa) architecture, an improved variant of BERT that removes the Next Sentence Prediction (NSP) task and enhances training efficiency through dynamic masking and large-batch optimization. WangchanBERTa was pre-trained on a large-scale Thai corpus of over 78 gigabytes, comprising diverse sources such as news articles, general web pages, social media content, and government documents, enabling rich contextual understanding of Thai linguistic structures across multiple domains.

Key differences between WangchanBERTa and standard BERT include: A SentencePiece tokenizer adapted for Thai word segmentation without relying on whitespace boundaries. A customized vocabulary covering Thai words, idioms, colloquialisms, and emojis common in user-generated content. RoBERTa-style training objectives that provide more stable and generalizable contextual embeddings.

WangchanBERTa is particularly well-suited for processing informal, mixed-sentiment Thai text, such as “ห้องเล็กแต่บริการดี” (small room but good service), where the model can distinguish both positive and negative sentiments within a single sentence. This aligns with findings by Lowphansirikul et al. [1], who reported WangchanBERTa’s superior contextual representations compared with multilingual Transformer models, and Suraratchai et al. [13], who demonstrated its strong performance in aspect-based sentiment analysis for Thai content.

Additional evidence for the effectiveness of Thai-specific Transformer architectures is provided by Harnmetta and Samanchuen [30], who demonstrated that transformer-based models, particularly Thai-specific architectures such as WangchanBERTa, outperform multilingual approaches in Thai sentiment classification tasks. Their findings further highlight the importance of language-specific contextual understanding in Thai sentiment analysis. Complementary studies in Thai sentiment research further indicate that deep contextual and LLM-assisted models can capture nuanced emotional patterns in large-scale social media data, emphasizing the need for Thai-focused Transformer architectures [22].

In summary, WangchanBERTa is a highly suitable Thai language model for hotel review analysis in this research, given its proven capability to process complex Thai expressions, interpret mixed sentiments, and support aspect-level understanding that aligns with the authentic linguistic characteristics of Thai user-generated content.

3.6 Performance Evaluation

3.6.1 Confusion Matrix

The Confusion Matrix is the primary tool for evaluating classification model performance, as it provides detailed insights into model predictions by comparing predicted values against actual values. It comprises four key components:

True Positive (TP): correctly predicted positive instances

True Negative (TN): correctly predicted negative instances

False Positive (FP): negative instances incorrectly predicted as positive

False Negative (FN): positive instances incorrectly predicted as negative

These four values form the foundation for calculating several important performance metrics used to evaluate model accuracy, detailed as follows [17], [22], [23], [24], [25].:

Accuracy

Accuracy measures the overall correctness of the model, reflecting the proportion of correct predictions among all samples [17], [22], [23], [24], [25].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision

Precision reflects the correctness of positive predictions. Higher precision indicates better reduction of False Positive errors [17], [22], [23], [24], [25].

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall

Recall measures the coverage of actual positive instances, particularly important when positive instances are minority class [17], [22], [23], [24], [25].

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score

F1-Score is the harmonic meaning of Precision and Recall, reflecting the balance between correctness and coverage [22], [23], [24], [25].

$$F - Measure = \frac{2 * Prc * Rcl}{Prc + Rcl} \quad (4)$$

3.6.2 Confidence Calibration

Confidence calibration assesses whether the model’s predicted probability scores accurately reflect the true correctness. A well-calibrated model should produce outcomes in which, for instance, predictions with a confidence of 0.8 are correct approximately 80% of the time.

Calibration is particularly important in real-world applications where decision-making relies on model uncertainty estimates. Poor calibration can lead to overconfident or underconfident predictions, potentially causing unreliable system behavior even when accuracy appears high.

3.6.3 Confidence Metrics

To quantify model confidence, several metrics derived from the model’s output layer are commonly used:

Max Score: The maximum softmax probability among all classes, indicating the model’s confidence in its top prediction. Higher values suggest stronger conviction in the predicted class [26], [27], [28].:

$$Max Score = \max_{i \in \{1, \dots, K\}} \left(\frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \right) \quad (5)$$

These metrics help analyze prediction uncertainty and identify low-confidence predictions that may require human review or further validation.

4. RESULTS AND DISCUSSION

4.1 Experimental Setup

Experiments were conducted on Google Colab using Python 3.10 with PyTorch, Transformers (HuggingFace), and Scikit-learn, on an NVIDIA Tesla T4 GPU (16 GB VRAM) and 12.7 GB RAM. The WangchanBERTa model was fine-tuned with a batch size of 16, a learning rate of $2e-5$, and the AdamW optimizer for 3-5 epochs.

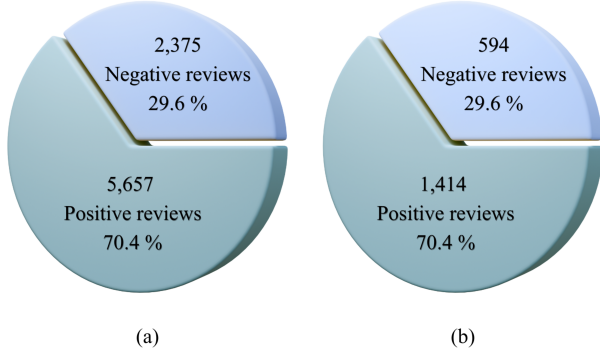


Fig. 2: Distribution of positive and negative sentiment classes in the training and test sets: (a) training set with 8,032 reviews; (b) test set with 2,008 reviews.

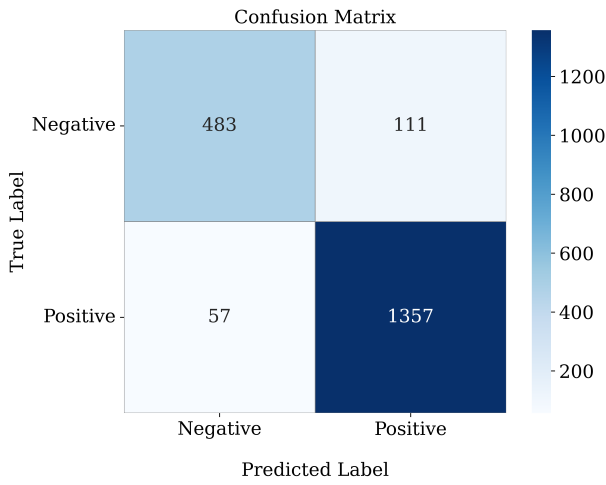


Fig. 3: Confusion matrix of WangchanBERTa model for sentiment classification.

4.2 Dataset Distribution and Sentiment Polarity

The data was partitioned using an 80:20 stratified split to maintain consistent class proportions across both subsets. As shown in Fig. 2(a), the training set contains 8,032 reviews with 5,657 positive reviews (70.4%) and 2,375 negative reviews (29.6%). The testing set in Fig. 2(b) comprises 2,008 reviews, consisting of 1,414 positive reviews (70.4%) and 594 negative reviews (29.6%).

The distribution reveals a consistent positive-to-negative ratio of approximately 7:3 across both sets, which is suitable for model training without requiring additional data rebalancing techniques.

4.3 Model Performance Evaluation

4.3.1 Overall Sentiment Classification

At the first level, the WangchanBERTa model was fine-tuned on the Thai hotel review dataset to classify text into two sentiment categories: Positive and Negative. The performance evaluation used four standard metrics: Precision, Recall, F1-score, and Accuracy.

Evaluation results on the test set demonstrate effective



Fig. 4: Word clouds of hotel reviews categorized by sentiment: (a) Positive reviews (b) Negative reviews.

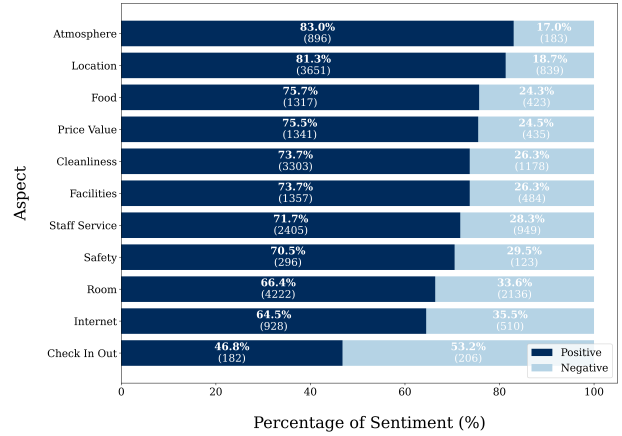


Fig. 5: Aspect-based sentiment distribution showing positive and negative guest opinions across service dimensions, predicted by fine-tuned WangchanBERTa.

sentiment classification, achieving a Precision of 91.55%, a Recall of 91.63%, a weighted F1-score of 91.52%, a macro F1-score of 89.69%, and an Accuracy of 91.63%. These results indicate the model's robust capability to capture contextual nuances and accurately classify sentiment in Thai review texts, even in cases involving complex sentence structures or implicit emotional expressions.

4.3.2 Aspect-Based Sentiment Classification

Following overall sentiment classification, the WangchanBERTa model was further fine-tuned for aspect-level sentiment analysis of hotel reviews across eleven service dimensions: Staff Service, Room, Cleanliness, Food, Location, Price & Value, Facilities, Check-in/Check-out Process, Internet, Safety, and Atmosphere. Performance evaluation used the same standard metrics as overall sentiment classification.

Table 2: Examples of Multi-Aspect Sentiment Detection Results.

No.	Review Message	Aspect Label	Sentiment (Confidence %)
1	The staff were very friendly, cheerful, and helpful.	Staff Service	Positive (96.52 %)
2	The room is spacious, but the furniture is getting old.	Room Facilities	Positive (95.64 %) Negative (97.16 %)
3	The property is very clean, but the bathroom has a slight musty odor.	Cleanliness Room	Positive (96.17 %) Negative (97.02 %)
4	Breakfast is delicious with plenty of choices, but the room is not clean	Food Room	Positive (95.88 %) Negative (98.02 %)
5	Great location, convenient for travel, close to the BTS.	Location	Positive (96.74 %)
6	The price is reasonable, but the Wi-Fi is very slow.	Price & Value Internet	Positive (92.41%) Negative (97.32 %)
7	Fast check-in with quick document handling.	Check-in / Check-out	Positive (98.12%)
8	The hotel has a nice atmosphere, but the parking lot is quite dark at night, making it feel a bit unsafe.	Atmosphere Safety	Positive (97.82%) Negative (96.42 %)

Experimental results on the test set demonstrate significantly improved performance, achieving a Precision of 93.62%, a Recall of 94.62%, an F1-score of 94.22%, and an Accuracy of 98.66%. These results indicate that the model effectively understands fine-grained review details and accurately identifies aspect-specific sentiments.

4.4 Confusion Matrix

The confusion matrix of the WangchanBERTa model for sentiment classification is shown in Fig. 3, with two categories: Positive and Negative. The matrix shows the number of correctly and incorrectly classified texts for each class, with darker colors indicating higher classification accuracy, reflecting the model's effectiveness at distinguishing different sentiments.

Analysis shows that the model achieves higher accuracy in classifying positive texts than negative ones. Misclassifications predominantly occur in two scenarios: (1) Mixed Sentiment — texts expressing both satisfaction and dissatisfaction within the duplicate content, and (2) Contrastive Expressions — statements containing contradictory elements within a single sentence.

Confusion matrix analysis also reveals patterns of systematic misclassification, such as negative reviews using polite language being incorrectly classified, and texts with complex sentence structures being misinterpreted contextually. These behavioral insights can inform model refinement to better understand Thai language "tone," thereby enhancing the accuracy of aspect-based insights extraction in subsequent phases, such as precisely identifying that users are "satisfied with service" but "dissatisfied with pricing".

Despite some misclassifications, the accuracy of 92.62% demonstrates that the model effectively learns and understands Thai review contexts at a level suitable for real-world application. Therefore, the confusion matrix in this research not only indicates model accuracy but also serves as a crucial tool for understanding how the model perceives human sentiment.

4.5 Data Analysis and Interpretation

The hotel review data analysis in this research employs a combined approach of textual analysis and aspect-based sentiment analysis to gain deep insights into customer opinions and emotions regarding their accommodation experiences across different service dimensions.

Fig. 4 illustrates the word cloud analysis of hotel reviews, highlighting the most frequently mentioned terms in both positive and negative contexts. The size of each word corresponds to its frequency of occurrence, enabling immediate identification of key themes and concerns expressed by guests across various aspects of their hotel experience.

Analysis reveals that frequently occurring terms in positive reviews reflect core strengths in cleanliness and service quality. Conversely, negative reviews predominantly relate to experiences with facilities and service efficiency. These visual insights not only enable hoteliers to grasp overall customer feedback quickly but also clearly identify areas requiring improvement and strengths to reinforce, aligning with the aspect-based sentiment analysis results presented in previous sections.

4.6 Aspect-based Insight Extraction Results

Following the training and performance evaluation of the fine-tuned WangchanBERTa model, this research applied the model to analyze customer feedback at the aspect level, yielding deep insights into accommodation experiences across different dimensions of Thai SME hotels.

From the analysis of 10,040 review texts, the system successfully detected keywords and classified opinions into 11 principal aspects as follows:

1. **Staff Service** – Employee service quality and hospitality
2. **Room** – Room quality and in-room comfort
3. **Cleanliness** – Cleanliness standards and environmental conditions

4. **Food** – Food and beverage offerings
5. **Location** – Hotel location and accessibility
6. **Price & Value** – Value for money
7. **Facilities** – Amenities (swimming pool, fitness center, parking, etc.)
8. **Check-in/Check-out Process** – Registration and departure procedures
9. **Internet** – Wi-Fi connectivity and internet speed
10. **Safety** – Security and safety measures
11. **Atmosphere** – Overall ambiance and interior design

To further interpret how guests perceive different service dimensions, the extracted aspect-level sentiments were aggregated and visualized to reveal the overall distribution of positive and negative opinions across the 11 hotel service aspects. This visualization highlights not only the volume of mentions within each aspect but also the polarity tendencies that reflect real customer experiences. Such distribution is essential for identifying strengths, diagnosing service pain points, and guiding managerial decision-making for Thai SME hotels. Fig. 5 presents the aspect-based sentiment distribution predicted by the fine-tuned WangchanBERTa model, illustrating how guests evaluated each service component in practical hotel operations.

4.7 Multi-Aspect Detection

The WangchanBERTa model can identify multiple aspects within a single review by segmenting the text into sub-sentences and evaluating aspects and sentiment polarity independently for each segment. This approach enables a precise understanding of which specific aspects elicit positive or negative responses from guests. Its sentiment orientation and confidence score accompany each detected aspect.

The aspect-based sentiment analysis results reveal customer satisfaction levels across different dimensions of the service experience, as illustrated in Table 2.

5. CONCLUSION

This paper presented a two-phase framework for Thai-language sentiment analysis in small and medium-sized hotel enterprises (SME hotels) using WangchanBERTa. The first phase focused on overall sentiment analysis, where WangchanBERTa was fine-tuned on 10,040 reviews collected from Agoda, Booking.com, Traveloka, and Trip.com. The fine-tuned model achieved an average accuracy of 91.63% and a macro F1-score of 89.69%, outperforming traditional classifiers and recurrent architectures such as LSTM and GRU. These results confirm the advantage of Transformer-based models in capturing long-range dependencies in Thai hotel reviews.

In the second phase, aspect-based sentiment analysis was employed to extract finer-grained insights. The system identified key service aspects, including price, cleanliness, staff service, and room comfort, and classified the sentiment polarity associated with each aspect. On the test set, the model obtained an accuracy of 91.63%,

with macro precision of 91.55%, recall of 91.63%, and a macro F1-score of 89.69%. The framework can also detect mixed opinions within a single review, enabling hotel operators to pinpoint specific strengths and weaknesses at the aspect level.

Future work includes expanding the dataset to additional platforms such as Google Reviews and Facebook, extending the framework to multilingual aspect analysis and exploring zero-shot or prompt-based learning to reduce annotation costs and adapt to emerging aspects. Another promising direction is integrating the proposed models with Business Intelligence dashboards, recommendation engines for real-time voice-of-customer monitoring and decision support. Overall, the proposed framework demonstrates that Thai-specific deep language models such as WangchanBERTa can substantially enhance sentiment analysis in the hospitality sector and can be generalized to related service domains.

ACKNOWLEDGEMENT

The authors thank the Center of Multidisciplinary Innovation Network Talent (MINT Center), Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai Campus, Thailand, for providing research facilities and resources.

AUTHOR CONTRIBUTIONS

The manuscript was written with contributions from all authors. They discussed the results and commented on the manuscript. All authors approved the final version of the manuscript.

REFERENCES

- [1] L. Lowphansirikul, C. Polpanumas, N. Jantrakulchai, and S. Nutanong, "WangchanBERTa: Pretraining transformer-based Thai language models," *arXiv preprint*, 2021, doi: 10.48550/arXiv.2101.09635.
- [2] T. Wolf et al., "Transformers: State-of-the-art natural language processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP)*, Online, Oct. 2020, pp. 38–45, doi: 10.18653/v1/2020.emnlp-demos.6.
- [3] C. Chantrapornchai and A. Tunsakul, "Information extraction tasks based on BERT and SpaCy on tourism domain," *ECTI Transactions on Computer and Information Technology*, vol. 15, no. 1, pp. 108–122, 2021, doi: 10.37936/ecti-cit.2021151.228621.
- [4] N. Khamphakdee and P. Seresangtakul, "Sentiment analysis for Thai language in hotel domain using machine learning algorithms," *Acta Informatica Pragensia*, vol. 10, no. 2, pp. 155–171, 2021, doi: 10.18267/j.aip.155.
- [5] N. Khamphakdee and P. Seresangtakul, "An Efficient Deep Learning for Thai Sentiment Analysis,"

- Data*, vol. 8, no. 5, art. 90, May 2023, doi: 10.3390/data8050090.
- [6] K. Namee, J. Polpinij, and B. Luaphol, "A hybrid approach for aspect-based sentiment analysis: A case study of hotel reviews," *Current Applied Science and Technology*, vol. 23, no. 2, 2023, doi: 10.55003/cast.2022.02.23.008.
 - [7] Y. Bai et al., "Is compound aspect-based sentiment analysis addressed by LLMs?," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, FL, USA, Nov. 2024, pp. 7836–7861, doi: 10.18653/v1/2024.findings-emnlp.460.
 - [8] T. Lapchaicharoenkit and P. Vateekul, "Machine reading comprehension using multi-passage BERT with Dice loss on Thai corpus," *ECTI Transactions on Computer and Information Technology*, vol. 16, no. 2, pp. 125–134, 2022, doi: 10.37936/ecti-cit.2022162.247799.
 - [9] Y.-C. Hua, Z. Zhou, Q. Han, Y. Li, R. He, and J. Zhou, "A systematic review of aspect-based sentiment analysis," *Artificial Intelligence Review*, vol. 57, art. 296, 2024, doi: 10.1007/s10462-024-10906-z.
 - [10] I. Nawawi, K. F. Ilmawan, M. R. Maarif, and M. Syafrudin, "Exploring tourist experience through online reviews using aspect-based sentiment analysis with zero-shot learning for hospitality service enhancement," *Information*, vol. 15, no. 8, art. 499, Aug. 2024, doi: 10.3390/info15080499.
 - [11] G. Kontonatsios, J. Clive, G. Harrison, T. Metcalfe, P. Sliwiak, H. Tahir, and A. Ghose, "An aspect-based sentiment analysis dataset of user reviews," *Neurocomputing*, vol. 562, art. 126867, 2023, doi: 10.1016/j.neucom.2023.126867.
 - [12] A. Ameer, S. Hamdi, and S. Ben Yahia, "Sentiment analysis for hotel reviews: A systematic literature review," in *Companion Proceedings of the 35th International Conference on Advanced Information Systems Engineering (CAiSE)*, Zaragoza, Spain, Jun. 2023, pp. 127–142, doi: 10.1145/3605152.
 - [13] K. Suraratchai and S. Phoomvuthisarn, "Thai language sentiment analysis with a hybrid method on WangchanBERTa-CNN-BiLSTM," *Journal of Information Science and Technology*, vol. 14, no. 2, pp. 1–11, 2024.
 - [14] W. Bunyatisai, S. Plubin, K. Jiamwattanapong, and B. Plubin, "Improving Thai sentiment analysis accuracy with emoji classification by deep learning and stacking models: A case study of hotel reviews," *Pakistan Journal of Life and Social Sciences*, vol. 23, no. 1, pp. 65–77, 2025, doi: 10.57239/PJLSS2025-23.1.007.
 - [15] N. Jensirisak, D. Tasachan, and P. Wanchai, "Aspect based sentiment analysis for Thai language from consumer reviews towards smartphone," *Journal of Science Ladkrabang*, vol. 34, no. 1, pp. 20–42, 2025.
 - [16] R. Wanasukapunt and S. Phimoltare, "Classification of abusive Thai language content in social media using deep learning," in *Proceedings of the 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2021, pp. 1–6, doi: 10.1109/JCSSE53117.2021.9503294.
 - [17] M. Nokkaew et al., "Analyzing online public opinion on Thailand–China high-speed train and Laos–China railway mega-projects using advanced machine learning for sentiment analysis," *Social Network Analysis and Mining*, vol. 14, no. 1, art. 15, 2024, doi: 10.1007/s13278-023-01168-8.
 - [18] Y. Wen, Y. Liang, and X. Zhu, "Sentiment analysis of hotel online reviews using the BERT model and ERNIE model—Data from China," *PLoS ONE*, vol. 18, no. 3, art. e0275382, Mar. 2023, doi: 10.1371/journal.pone.0275382.
 - [19] M. Thomson, H. Murfi, and G. Ardaneswari, "BERT based hybrid deep learning with text augmentation for sentiment analysis of Indonesian hotel reviews," in *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, Rome, Italy, Jul. 2023, pp. 468–473, doi: 10.5220/0012127400003541.
 - [20] S. C. Sahin and C. Eyupoglu, "Aspect-based sentiment analysis for hospitality industry applications: A systematic literature review," *Applied Computer Systems*, vol. 30, no. 1, pp. 53–67, 2025, doi: 10.2478/acss-2025-0007.
 - [21] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, "Leveraging large language models in tourism: A comparative study of the latest GPT Omni models and BERT NLP for customer review classification and sentiment analysis," *Information*, vol. 15, no. 12, art. 792, 2024, doi: 10.3390/info15120792.
 - [22] M. Nokkaew et al., "Hidden emotional trends on social media regarding the Thailand–China high-speed railway project: A deep learning approach with ChatGPT integration," *Social Network Analysis and Mining*, vol. 14, no. 1, art. 175, 2024, doi: 10.1007/s13278-024-01340-8.
 - [23] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing aspect-based sentiment analysis using BERT," *Applied Sciences*, vol. 15, no. 1, art. 172, Jan. 2025, doi: 10.3390/app15010172.
 - [24] I. A. Kandhro, F. Ali, M. Uddin, A. Kehar, and S. Manickam, "Exploring aspect-based sentiment analysis: An in-depth review of current methods and prospects for advancement," *Knowledge and Information Systems*, vol. 66, pp. 3639–3669, 2024, doi: 10.1007/s10115-024-02104-8.
 - [25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Minneapolis, MN, USA, Jun. 2019, pp. 4171–

4186, doi: 10.18653/v1/N19-1423.

- [26] J. Geng, F. Cai, Y. Wang, H. Koeppel, P. Nakov, and I. Gurevych, "A survey of confidence estimation and calibration in large language models," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Mexico City, Mexico, Jun. 2024, pp. 6577–6595, doi: 10.18653/v1/2024.naacl-long.366.
- [27] Y. Cao, H. Mozannar, L. Feng, H. Wei, and B. An, "In defense of softmax parametrization for calibrated and consistent learning to defer," in *Proceedings of the 37th Annual Conference on Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, USA, Dec. 2023, pp. 38485–38503.
- [28] F. Adjed, S. Herbin, and A. LeCoz, "Confidence calibration of classifiers with many classes," in *Proceedings of the 38th Annual Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, Dec. 2024, doi: 10.52202/079017-2469.
- [29] W. Phatthiyaphaibun et al., "PyThaiNLP: Thai natural language processing in Python," in *Proceedings of the 3rd Workshop on Natural Language Processing Open Source Software (NLP-OSS 2023)*, Singapore, Dec. 2023, pp. 21–30, doi: 10.18653/v1/2023.nlposs-1.4.
- [30] P. Harnmetta and T. Samanchuen, "Sentiment Analysis of Thai Stock Reviews Using Transformer Models," in *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Chiang Mai, Thailand, 2022, pp. 1-6, doi: 10.1109/JCSSE54846.2022.9960246.



Thanachok Suwan is currently pursuing the B.Sc. degree in Computer Science and Information at Khon Kaen University, Nong Khai Campus, Thailand. His research interests include artificial intelligence, natural language processing, sentiment analysis, and deep learning. His current research focuses on aspect-level sentiment analysis for Thai hotel reviews using WangchanBERTa to support data-driven decision-making.



Manussawee Nokkaew received the Ph.D. degree in Electrical and Computer Engineering from Khon Kaen University, Thailand, in 2024. From 2015 to 2025, he worked as an analyst and consultant at Software Park Thailand under the National Science and Technology Development Agency (NSTDA). During this period, his work focused on managing projects to empower software and digital entrepreneurs in accessing domestic and international markets, and driving digital transformation among industrial enterprises. Currently, he is a Lecturer in the Computer and Information Science Program, Technology and Engineering Sector, at Khon Kaen University, Nong Khai Campus. His research interests include artificial intelligence (AI), particularly generative AI and natural language processing (NLP), as well as machine learning, deep learning, big data analytics, and sustainability.



Chayada Surawanitkun received the B.Eng. degree in Electrical Engineering (1st class honor) and the Ph.D. degree in Electrical Engineering from Khon Kaen University, Thailand, in 2009 and 2013, respectively. She was granted the B.Eng. Scholarship from the Industrial/University Cooperative Research Center in HDD Component, supported by the Ministry of Science and Technology, and a Ph.D. scholarship from the Royal Golden Jubilee Ph.D. Program under The Thailand Research Fund. She is Program Chair of Master's and Doctoral programs in Electrical and Computer Engineering of the Faculty of Interdisciplinary Studies, Department of Technology and Engineering, Khon Kaen University, Nong Khai Campus, Thailand. Her research interests include power systems, Smart IoTs, data storage technology, magnetic random-access memory, thermal analysis of nanodevices, nanotechnology, renewable and sustainable energy, and energy management systems.



mender systems, and sentiment analysis.

Kanda Sorn-In is an Assistant Professor in the Computer and Information Science Program, Department of Technology and Engineering, Faculty of Interdisciplinary Studies, Khon Kaen University, Thailand. She received her B.Sc. and M.Sc. degrees in Computer Science and her Ph.D. in Information Studies from Khon Kaen University. Her research interests include data analysis, information systems design, artificial intelligence, neural networks, business intelligence, recommender systems, and sentiment analysis.



Nongram Mueanrit received her Ph.D. in Information Studies from Khon Kaen University, Thailand, in 2013. She is currently serving as an Assistant Professor in the Department of Technology and Engineering, Faculty of Interdisciplinary Studies, Khon Kaen University. Her research interests encompass a broad spectrum of interdisciplinary domains, including information behavior analytics, consumer behavior analysis, user experience design, business data analytics, and smart tourism systems. She has contributed to the academic community through publications in internationally recognized peer-reviewed journals, including *Energy*, among others.



Kwankamol Nongpong currently serves as the Chairperson of the Department of Computer Science at the Vincent Mary School of Engineering, Science and Technology, Assumption University, Thailand. Her research interests encompass text processing, natural language processing, big data analytics, program analysis, and enterprise applications. In addition to her academic responsibilities, she maintains active engagement with industry through consultancy work for both private sector organizations and government agencies. Her areas of expertise include market research, data standardization, project management, business process improvement, and enterprise resource planning (ERP) system development.



Tapanan Yeophantong is currently a full-time lecturer at the Department of Design and Digital Innovation, Martin de Tours School of Management and Economics, Assumption University. He is also a co-founder and managing director at Praram Nine Technology, a digital transformation solution provider specialising in AI technologies for smart industry, large-scale facilities, and digital marketing platforms. With more than 20 years of experience in artificial intelligence,

his research expertises include multi-agent systems, machine/deep learning, computer vision, natural language processing, and complex event processing.



Wisut Supasai received the Ph.D. degree in electrical and computer engineering from Khon Kaen University, Thailand, in 2024. Currently, He holds the position of researcher at the Faculty of Science, Ramkhamhaeng University. His research interests are computational spintronics, magnetic storage, magnetic random-access memory (MRAM), photonic computing, data science, AI technology and electric vehicle and renewable energy.



Apirat Siritaratiwat received the B.Eng. degree in Electrical Engineering from Khon Kaen University, Thailand, in 1992. After working in industry for a few years, he joined the Department of Electrical Engineering, Khon Kaen University (KKU) in 1994 and continued his postgraduate study at the University of Manchester, U.K., between 1995 and 1999. He has done research in HDD and is one of the pioneer researchers of ESD/EOS and EMI in recording heads with the KKU-

Seagate Cooperation Research Laboratory, Faculty of Engineering, Khon Kaen University. Currently, He works at Department of IoT and Information Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang, Bangkok. His research interests are ESD/EOS, EMI, renewable energy and energy management system.