

MAHASAMUT: AI-Powered Thai Tourism using Multimodal Agents

Nongnuch Ketui^{1,*}, Parinthapat Pengpun², Konthee Boonmeeprakob³, Pitikorn Khlaisamniang⁴ and Thepchai Supnithi⁵

¹Rajamangala University of Technology Lanna

²Bangkok Christian International School

³Big Data Institute

⁴Artificial Intelligence Association of Thailand

⁵NECTEC

Abstract

This paper presents "MAHASAMUT" an application designed to enhance the travel experience in Thailand. MAHASAMUT stands for Multimodal Assistant for Helping Adventurous Sightseers and All Manner of Unique Travels. The system leverages AI technologies to provide comprehensive support for travelers. It features an ASR (Automatic Speech Recognition) system capable of understanding various Thai dialects and a TTS (Text-to-Speech) module for seamless communication with local people. Additionally, it includes a VQA (Visual Question Answering) model that generates descriptive captions from images, making it easier for travelers to comprehend their surroundings. The VQA model is also able to perform OCR (Optical Character Recognition) to identify and translate Thai text into English, aiding travelers in navigating signs, menus, and other written materials. Furthermore, MAHASAMUT can generate images from textual descriptions, such as visualizing a dish from a menu that lacks images. By integrating these multimodal capabilities, MAHASAMUT enhances the accessibility and enjoyment of traveling in Thailand, providing an intuitive and interactive guide for tourists.

Keywords

Multi-Agent AI Systems, Multimodality, AI Tourism,

1. Introduction

The tourism industry, a cornerstone of many economies, faces increasing challenges in the era of globalization and digital transformation. Nowhere is this more evident than in Thailand, where tourism contributes significantly to the GDP [1, 2]. However, traditional tourism models often lead to issues such as overcrowding at popular destinations, cultural misunderstandings, and economic disparities that can negatively impact local communities [3].

Artificial Intelligence (AI) has emerged as a powerful tool to address these challenges, offering potential solutions for personalized experiences, language barriers, and sustainable tourism practices [4, 5]. Recent advancements in natural language processing, computer vision, and multi-agent systems have opened new avenues for enhancing the tourism experience [6, 7]. However, existing AI applications in tourism often focus on narrow aspects such as recommendation systems or language translation, without fully addressing the comprehensive nature of cultural tourism [5, 8, 9].

There is a growing need for comprehensive AI solutions that can navigate the intricate landscape of cultural tourism, particularly in linguistically and culturally diverse countries like Thailand. Such solutions must not only provide practical assistance to tourists but also promote sustainable practices and support local economies [10].

In this paper, we present MAHASAMUT, an AI-powered multi-agent system designed to enhance cultural tourism in Thailand. Our system integrates a multitude of AI technologies to combine into a comprehensive system, including:

The 9th Linguistic and Cognitive Approaches to Dialog Agents Workshop, November 19, 2024, Kyoto, Japan

*Corresponding Author.

✉ nongnuchketui@rmutl.ac.th (N. Ketui)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. A **culturally-aware large language model** (LLM) fine-tuned for tourism contexts integrated with retrieval augmented generation (RAG) over the internet
2. An **automatic speech recognition** (ASR) system capable of understanding Thai dialects
3. A **visual question-answering** (VQA) component for landmark and artifact interpretation
4. **Image and speech generation** components for easy two-way communication
5. A coordinated **multi-agent architecture** for complex query resolution

MAHASAMUT aims to provide personalized, culturally sensitive guidance to tourists while promoting off-the-beaten-path experiences and supporting local economies. By leveraging AI to bridge language and cultural gaps, our system facilitates more meaningful interactions between tourists and local communities.

The rest of this paper is organized as follows: Section 2 reviews related work in AI applications for tourism. Section 3 describes the system architecture and key components of MAHASAMUT. Section 4 details the application of our system in cultural tourism scenarios. We discuss the implications and future directions of our work in Section 5, before concluding in Section 6.

2. Related Work

AI Applications in Tourism The travel and tourism industry has increasingly adopted AI technologies to enhance various aspects of the tourist experience. For instance, Filieri et al. (2020) explored the characteristics of AI start-ups in Europe that focus on tourism, highlighting the significance of AI in reshaping the industry through applications in big data, machine learning, and natural language processing. These technologies enable marketing automation, segmentation, and customization, significantly benefiting the tourism supply chain, particularly in the pre-trip and post-trip phases [11]. Additionally, the development of chatbots for smart tourism mobile applications has shown promise in improving tourist experiences by providing tailored information and assistance before, during, and after their visits [12].

Supporting LLM, ASR, and TTS

The integration of automatic speech recognition (ASR) and text-to-speech (TTS) systems in tourism applications has been explored to facilitate seamless communication between tourists and locals. For instance, the development of speech-to-speech translation interfaces has proven effective in enabling tourists to interact using their native languages, thereby enhancing their travel experiences [13]. Furthermore, visual question answering (VQA) systems, which can handle multilingual and code-mixed queries, are increasingly being used to provide detailed information about tourism objects based on images captured by mobile devices [14]. These technologies collectively contribute to a more interactive and accessible travel experience, aligning with the goals of MAHASAMUT.

3. Methodology

3.1. System Overview

The MAHASAMUT system is a multi-agent AI architecture designed to enhance cultural tourism experiences in Thailand. At its core, the system leverages a series of specialized agents orchestrated by the PhuKhao LLM. It first analyzes user queries and determines which specialized components should be employed to address the request effectively, allowing for dynamic and context-aware processing of user inputs.

PhuKhao LLM, a large language model fine-tuned on a proprietary synthetic dataset, forms the backbone of the system. It has been trained on a diverse range of tasks and topics, including cultural content specific to Thai tourism, enabling it to handle a wide array of queries with cultural sensitivity and accuracy.

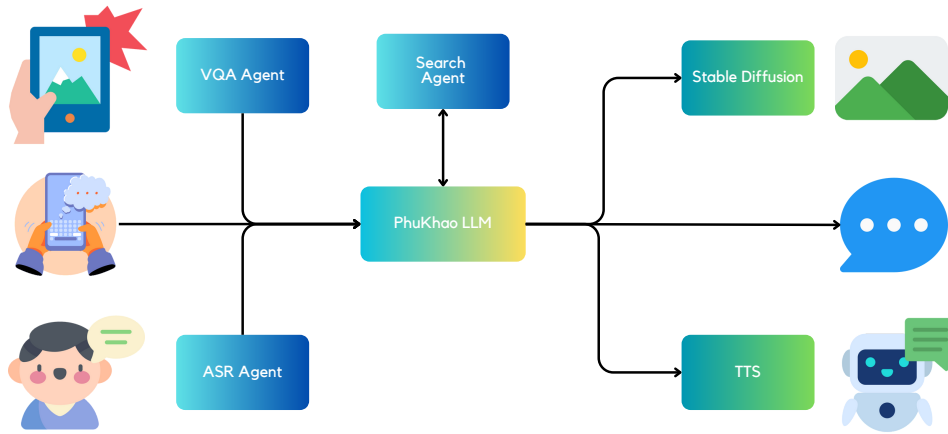


Figure 1: Diagram of the MAHASAMUT System

MAHASAMUT incorporates several key components that work in concert to provide comprehensive assistance to users. These include a Search Agent utilizing LangChain [15] and DuckDuckGo ¹ for retrieving and summarizing up-to-date information, a Visual Question Answering (VQA) Agent based on a fine-tuned version of PaliGemma [16] for image-related queries, and an Image Generation Agent that creates contextually relevant images using Stable Diffusion 3 [17].

To facilitate seamless interaction in various languages and modalities, the system also includes specialized modules for Automatic Speech Recognition (ASR), Text-to-Speech (TTS), and translation. These components are crucial for navigating the diverse linguistic landscape of tourism in Thailand. The typical workflow of the MAHASAMUT system begins with user input, which is then analyzed by the Judgment Agent. Based on this analysis, the system engages the appropriate specialized agents or modules for processing. The core PhuKhao LLM then processes the query, incorporating any additional context from the specialized components. Finally, the system generates a response, which may include text, images, or speech, depending on the nature of the query and the most appropriate format for the user.

3.2. PhuKhaoLLM

The PhuKhao LLM is the core language model of the MAHASAMUT system, designed to handle Thai cultural and tourism contexts. It is built upon Typhoon-1.5 [18], a model fine-tuned by SCB10x based on the LLaMA 3 8B [19] architecture. Key features of PhuKhao LLM include:

- 8 billion parameters
- Fine-tuned using QLoRa [20]
- Training data: approximately 20,000 examples covering various tourism-related domains (generated using Seed-Free Synthetic Instruct [21])
- Task types: summarization, closed question-answering, open-ended conversation, and multiple-choice questions
- Fine-tuned for one epoch to balance knowledge and general capabilities

The model is tailored to handle queries related to Thai cuisine, local customs, historical sites, and travel information. While there are no standardized benchmarks for Thai tourism-specific language models, internal evaluations show improvements over the base Typhoon model in tourism-related tasks. PhuKhao LLM’s primary focus is on natural language understanding and generation within the context of Thai culture and tourism, aligning with the MAHASAMUT system’s objectives.

¹<https://duckduckgo.com/>

Table 1
Performance of VQA Models

Model	BLEU Score
(git-base)-PyThaiNLP	49.8
CLIP_laion2B-HoogBERTa	50.6
(blip2-opt-2.7b-coco)-PyThaiNLP	52.0
DinoV2-HoogBERTa	52.1
(blip2-flan-t5-xxl)-PyThaiNLP	53.2
Idefics2	54.5
PaliGemma	61.5

3.3. VQA Agent

To develop a robust Visual Question Answering (VQA) agent within the MAHASAMUT application, we employed the PaliGemma model family, specifically focusing on the pretrained (PT) models. These models are particularly suited for transfer learning, making them ideal for fine-tuning on tasks such as image captioning and VQA in the Thai language. We initially tested the mixed (mix) models due to their diverse training on various tasks, which provided valuable insights into their performance across different scenarios. For fine-tuning, we selected models with a 448x448 resolution and float32 precision, balancing memory efficiency and performance to capture and interpret image nuances accurately.

The training process utilized the MSCoco [22] and IPU2024 [23] datasets. The MSCoco dataset, a benchmark for image captioning, included 118,287 training images, 5,000 validation images, and 40,670 test images, offering a diverse set of images with corresponding captions. The IPU2024 dataset, tailored to Thai cultural contexts, included images related to Thai food and travel destinations, enhancing the model's ability to generate accurate Thai captions. This domain-specific dataset was crucial for providing localized and contextually relevant support to users.

Fine-tuning the captioning models yielded exceptional results, particularly with the PaliGemma model, which achieved the highest BLEU score of 61.5 (see Table 1). This score indicates its superior capability in generating accurate Thai captions tasks. The BLEU score is a metric for evaluating the quality of text generated by a model, with higher scores reflecting better performance. The models using PyThaiNLP leverage English captioning translated into Thai, while models like CLIP-HoogBERTa and DinoV2-HoogBERTa employ encoder-decoder architectures for direct Thai captioning. The superior performance of PaliGemma underscores its effectiveness and reliability in interpreting about images.

Fine-tuning the models yielded exceptional results, with the system excelling in generating Thai captions, performing VQA tasks, and recognizing Thai text through Optical Character Recognition (OCR). The model's VQA capabilities allowed it to interpret and answer questions about images, making the travel experience more engaging and informative. Additionally, the OCR functionality enabled the translation of Thai text from images, such as signs and menus, into English, aiding non-Thai speaking travelers in navigating their surroundings. These advancements underscore the potential of AI in transforming the tourism industry by offering personalized and contextually relevant assistance, enhancing the travel experience in Thailand.

3.4. ASR Agent

To enable MAHASAMUT to understand spoken Thai, including various dialects, we implemented an Automatic Speech Recognition (ASR) system leveraging state-of-the-art deep learning models and techniques specifically adapted for the Thai language. We selected the Whisper [24] model family, particularly the `openai/whisper-large-v3`, due to its robust performance across multiple languages and its ability to handle diverse accents and acoustic conditions.

The fine-tuning process involved adapting the Whisper model to Thai-specific phonetics and language patterns using transfer learning techniques. This process focused on improving accuracy for Thai

Table 2
Performance of ASR Models

Model	WER Score
wav2vec2-large-xlsr-53	0.199
whisper-medium	0.163
whisper-large-v3	0.130

pronunciation, tonal distinctions, and dialect variations. The dataset used for fine-tuning the ASR system is from the Thai Dialects Speech Recognition Challenge [25], consisting of speech samples from four regional dialects: Central (CT), Eastern (ET), Northern (NT), and Southern (ST). This dataset is divided into training, development, and test sets, with balanced distributions across dialects. Specifically, it includes 48,000 utterances from 480 speakers in the training set, 5,200 utterances from 52 speakers in the development set, and 2,800 utterances from 28 speakers in the test set.

For the ASR agent, we utilized two models: the Thai dialects fine-tuned Whisper for Thai ASR and the original Whisper for English ASR. Before processing the input, a classification step determines whether the input is in Thai or English, ensuring that the appropriate model is used. This dual-model approach allows MAHASAMUT to accurately interpret and respond to spoken language, enhancing communication for travelers in Thailand by providing seamless interaction across multiple dialects and languages.

The fine-tuning results demonstrated improvements in the model’s ability to accurately transcribe Thai speech. As shown in Table 2, the Whisper-large-v3 model achieved the best Word Error Rate (WER) of 0.130. WER is a common metric for evaluating the performance of an ASR system, representing the percentage of words that were incorrectly predicted. A lower WER indicates higher accuracy. The superior performance of the Whisper-large-v3 model in recognizing Thai speech, especially considering the diverse dialects and tonal variations, underscores its effectiveness and reliability. This performance solidified its selection as the primary ASR component for MAHASAMUT.

3.5. Search Engine Agent

To enhance MAHASAMUT’s information retrieval capabilities, we integrated a search engine component using the DuckDuckGo search API. This integration enables the system to access up-to-date information from the web, providing users with relevant and current data about their travel queries. The search engine integration process involves several steps to ensure efficient and accurate information retrieval.

The first step involves query formulation and search execution. We implemented the DuckDuckGo search API to perform web searches based on user queries. A query formulation mechanism was developed to translate user input into effective search terms, considering Thai language nuances and travel-specific contexts. The search engine is configured to retrieve the top 5 results for each query, balancing comprehensive coverage with processing efficiency. This approach ensures that users receive the most relevant and timely information for their travel needs.

Next, we focus on web scraping and content extraction. Utilizing the Beautiful Soup library ², we parse the HTML content of each retrieved URL, specifically extracting text contained within ‘<p>’ tags, which typically hold the main body content of web pages. We implemented error handling and timeout mechanisms to manage potential issues with webpage access or parsing, ensuring robust and reliable content extraction.

Finally, content summarization and context aggregation are performed. Then, PhuKhao LLM is employed to summarize the extracted content from each of the top 5 search results. We developed prompts for the LLM to produce concise, relevant summaries focused on travel-related information. Each search result is processed independently, generating five separate summaries. These summaries are then

²<https://pypi.org/project/beautifulsoup4/>

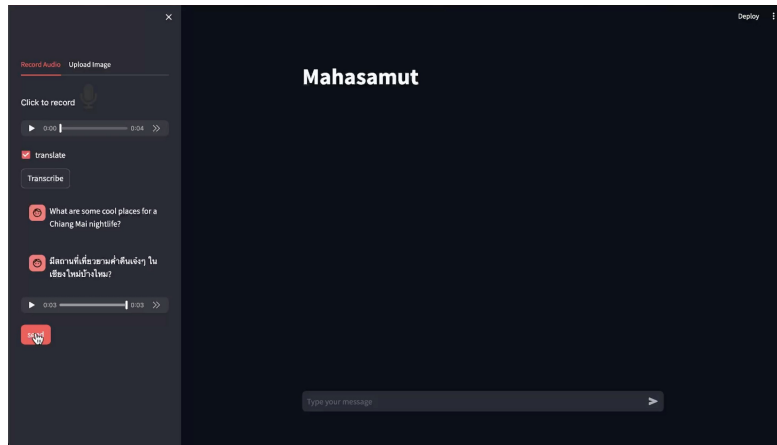


Figure 2: ASR Model visible in the User Interface

concatenated to create a comprehensive context. Another PhuKhaoLLM instance, generates answers to user questions based on this aggregated context. This multi-step process ensures that MAHASAMUT can provide accurate and detailed answers to user queries, enhancing the travel experience with timely and relevant information.

3.6. Image Generation Agent

The Image Generation Agent within MAHASAMUT utilizes the advanced capabilities of Stable Diffusion 3 (SD3) to produce high-quality, contextually relevant images from text prompts. Stable Diffusion 3, developed by Stability AI, employs a sophisticated Multimodal Diffusion Transformer (MMDiT) architecture that uses separate sets of weights for text and image data, allowing for more accurate and coherent image generation.

To generate prompts for SD3, the Image Generation Agent relies on the PhuKhao LLM. This LLM processes user inputs, which can be in either Thai or English, and generates comprehensive prompts in English that are suitable for image generation. By incorporating few-shot examples of Thai menu dishes, the LLM is able to understand and accurately reflect cultural nuances in its prompts. For instance, when a user describes a dish like "Pad Thai", the LLM generates a prompt such as: "A detailed image of Pad Thai, a popular Thai stir-fried noodle dish, garnished with shrimp, peanuts, and lime, served on a traditional Thai plate"

Once the LLM generates a prompt, it is fed into the Stable Diffusion 3 model, which then produces a high-quality image that closely matches the description. This process enhances various aspects of the travel experience, such as visualizing dishes from menus, creating images of landmarks or cultural artifacts, and providing visual navigation aids.

4. User Interaction

4.1. Two-way Communication

As shown in Figure 2, users can easily speak into MAHASAMUT. The user's voice will be interpreted by our English ASR Agent. Users can also speak Thai directly and our ASR Agent and it can be interpreted and understood as well. Utilizing the AI For Thai API for our Translation and TTS module, MAHASAMUT is also able to directly translate and speak the user's query— enabling two-way interaction between the users and a native speaker.

Mahasamut

👋 Hello, I need help planning a one-day trip to Chiang Mai, Thailand. Please include estimated times for each activity, brief descriptions, and any additional tips or recommendations. Please search on the internet and attached the references. Thank you!

📅 Planning a one-day trip to Chiang Mai, Thailand sounds like an exciting adventure! Based on the provided context and online research, here's a suggested itinerary for you:

9:00 AM - 10:00 AM: Visit Doi Suthep
Start your day by visiting Wat Phra That Doi Suthep, a iconic temple located on a hill overlooking the city. Take a 30-minute drive up the mountain and enjoy the stunning views of Chiang Mai. You can also explore the temple grounds and learn about its history.

10:30 AM - 12:00 PM: Explore the Old City
After visiting Doi Suthep, head back down to the Old City and explore its historic streets. Visit Wat Phra Singh, a beautiful temple located at the end of Rachadamnoen Road, and explore the surrounding markets and shops.

12:00 PM - 1:30 PM: Lunch at a local restaurant
Take a break and enjoy lunch at a local restaurant. Try some traditional Thai dishes like pad thai, tom yum soup, or green curry.

1:30 PM - 3:30 PM: Visit the Night Bazaar
In the afternoon, head to the Chiang Mai Night Bazaar, a bustling market filled with street food, souvenirs, and local handicrafts. You can find everything from traditional Thai clothing to handmade jewelry and ceramics.

In the evening, visit the Chiang Mai City Arts & Cultural Center, a museum showcasing the city's rich cultural heritage. You can see exhibits on the history of Chiang Mai, as well as traditional Thai art and crafts.

7:00 PM - 9:00 PM: Dinner at a local restaurant
End your day with dinner at a local restaurant. Try some street food or visit a restaurant that serves traditional Thai cuisine.

Additional tips and recommendations:

- Wear comfortable shoes and clothing, as you'll be doing a lot of walking and exploring.
- Bring sunscreen and a hat to protect yourself from the sun.
- Try to avoid visiting during peak hours (11:00 AM - 2:00 PM) to avoid the crowds.
- Be respectful of local customs and traditions, especially when visiting temples.
- Don't forget to try some of the local street food, which is delicious and affordable.

References:

- <https://www.onedaytravels.com/one-day-in-chiang-mai-thailand/>
- <https://www.tourismthailand.org/en/chiang-mai/chiang-mai>
- <https://www.onedaytravels.com/one-day-in-chiang-mai-thailand/>
- <https://www.onedaytravels.com/one-day-in-chiang-mai-thailand/>
- <https://www.onedaytravels.com/one-day-in-chiang-mai-thailand/>
- <https://www.onedaytravels.com/one-day-in-chiang-mai-thailand/>

I hope this helps you plan your one-day trip to Chiang Mai, Thailand!

Figure 3: MAHASAMUT planning a trip using the PhuKhaosLLM and the Search Agent

4.2. Trip Planning

As shown in Figure 3, MAHASAMUT is able to utilize its searching capabilities to provide users with up-to-date information, as well as citations of the information too. By providing MAHASAMUT with a query, it is able to output a detailed trip itinerary as shown. Furthermore, it is able to suggest local restaurants and also iconic places as well.

4.3. Visual Capabilities



Figure 4: An example of MAHASAMUT doing OCR on Thai Text

As shown in Figure 4, MAHASAMUT can directly read the menu which is written in Thai. Users can also then proceed to ask MAHASAMUT to generate an image of the menu if they are unsure what it is as shown in Figure 5. MAHASAMUT can also logically reason over the food— providing information such as the spiciness of the dish, it's composition, normal price, its history, etc.

5. Discussion

The development and implementation of MAHASAMUT present several implications for cultural tourism in Thailand and potentially for AI applications in tourism more broadly.

MAHASAMUT has the potential to greatly enhance the tourist experience in Thailand by providing personalized, culturally sensitive guidance. By leveraging AI to bridge language barriers and offer deep cultural insights, the system could facilitate more meaningful interactions between tourists and local

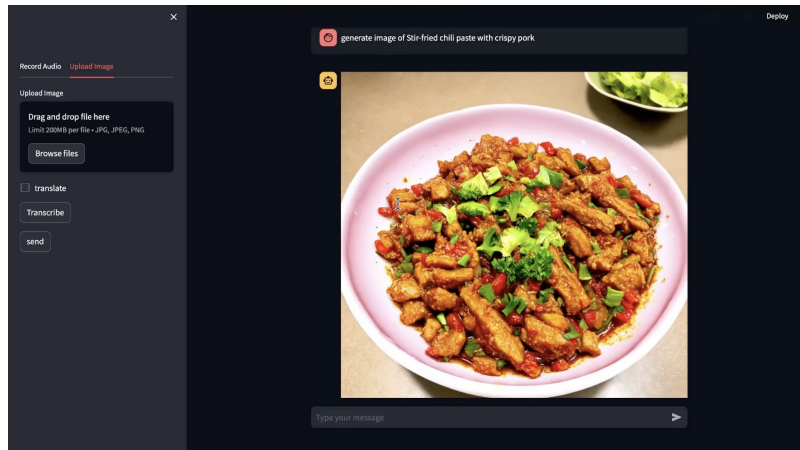


Figure 5: MAHASAMUT generating an image of a Thai dish

communities. This may lead to increased cultural understanding, more diverse tourism experiences, and potentially, economic benefits for less-visited areas.

Future development of MAHASAMUT will focus on improvements across all components. This includes enhancing the accuracy and breadth of information provided, further optimizing system performance, and expanding language support. A key area for improvement is the user interface, which needs refinement to ensure accessibility and ease of use for a diverse range of users.

6. Conclusion

This paper presented MAHASAMUT, an AI-powered multi-agent system designed to enhance cultural tourism in Thailand. By integrating the PhuKhao LLM, visual question answering, speech recognition, and image generation capabilities, MAHASAMUT addresses key challenges in cultural tourism such as language barriers and personalized recommendations. Our system demonstrates the potential of combining multiple AI technologies to create a comprehensive tourism assistance tool. While initial results are promising, further refinement of the user interface and extensive real-world testing are needed. MAHASAMUT represents a step towards more immersive and culturally sensitive AI applications in tourism. As development continues, we aim to balance technological innovation with ethical considerations, potentially contributing to more enriching and sustainable travel experiences.

Acknowledgments

We would like to express our gratitude to the Artificial Intelligence Association of Thailand (AIAT) for providing the essential facilities and resources that made this research possible. This work was financially supported by the National Science, Research and Innovation Fund (NSRF) through the Program Management Unit for Human Resources and Institutional Development, Research and Innovation [Grant Number: B13F670080] under project "Development of High Caliber Manpower in Artificial Intelligence and Prompt Engineer for Industry Support, focusing on Health, Energy and Environment, Finance and Digital Industry".

References

- [1] C. Theparat, Prayut: Zones vital for growth, Bangkok Post (2019). URL: <https://www.bangkokpost.com/business/1753349/prayut-zones-vital-for-growth>, accessed online.
- [2] World Travel & Tourism Council, Travel & tourism economic impact 2023: Thailand (2023). URL: <https://researchhub.wttc.org>, accessed online.

- [3] R. Dodds, R. Butler, The phenomena of overtourism: a review, *International Journal of Tourism Cities ahead-of-print* (2019). doi:10.1108/IJTC-06-2019-0090.
- [4] U. Gretzel, M. Sigala, Z. Xiang, C. Koo, Smart tourism: foundations and developments, *Electronic Markets* 25 (2015) 179–188. URL: <https://doi.org/10.1007/s12525-015-0196-8>. doi:10.1007/s12525-015-0196-8.
- [5] M. Li, D. Yin, H. Qiu, B. Bai, A systematic review of ai technology-based service encounters: Implications for hospitality and tourism operations, *International Journal of Hospitality Management* 95 (2021) 102930. URL: <https://www.sciencedirect.com/science/article/pii/S0278431921000736>. doi:<https://doi.org/10.1016/j.ijhm.2021.102930>.
- [6] D. Suhartanto, A. Brien, N. Sumarjan, N. Wibisono, Examining attraction loyalty formation in creative tourism, *International Journal of Quality and Service Sciences* 10 (2018) 163–175. URL: <https://doi.org/10.1108/IJQSS-08-2017-0068>. doi:10.1108/IJQSS-08-2017-0068.
- [7] D. Buhalis, Y. Sinarta, Real-time co-creation and nowness service: lessons from tourism and hospitality, *Journal of Travel Tourism Marketing* 36 (2019) 563–582. URL: <https://doi.org/10.1080/10548408.2019.1592059>. doi:10.1080/10548408.2019.1592059.
- [8] J. Borràs, A. Moreno, A. Valls, Intelligent tourism recommender systems: A survey, *Expert Systems with Applications* 41 (2014) 7370–7389. URL: <https://www.sciencedirect.com/science/article/pii/S0957417414003431>. doi:<https://doi.org/10.1016/j.eswa.2014.06.007>.
- [9] Y. Yu, Application of translation technology based on ai in translation teaching, *Systems and Soft Computing* 6 (2024) 200072. URL: <https://www.sciencedirect.com/science/article/pii/S2772941924000012>. doi:<https://doi.org/10.1016/j.sasc.2024.200072>.
- [10] X. Font, S. McCabe, Sustainability and marketing in tourism: its contexts, paradoxes, approaches, challenges and potential, *Journal of Sustainable Tourism* 25 (2017) 869–883. URL: <https://doi.org/10.1080/09669582.2017.1301721>. doi:10.1080/09669582.2017.1301721.
- [11] R. Filieri, E. D’Amico, A. Destefanis, E. Paolucci, E. Raguseo, Artificial intelligence (ai) for tourism: an european-based study on successful ai tourism start-ups, *International Journal of Contemporary Hospitality Management ahead-of-print* (2021). doi:10.1108/IJCHM-02-2021-0220.
- [12] L. Benaddi, C. Ouaddi, A. Jakimi, B. Ouchao, Towards a software factory for developing the chatbots in smart tourism mobile applications, *Procedia Computer Science* 231 (2024) 275–280. URL: <https://www.sciencedirect.com/science/article/pii/S1877050923022159>. doi:<https://doi.org/10.1016/j.procs.2023.12.203>, 14th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 13th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (EUSPN/ICTH 2023).
- [13] M. Cettolo, A. Corazza, G. Lazzari, F. Pianesi, E. Pianta, L. M. Tovenà, A speech-to-speech translation based interface for tourism, in: D. Buhalis, W. Schertler (Eds.), *Information and Communication Technologies in Tourism 1999*, Springer Vienna, Vienna, 1999, pp. 191–200.
- [14] D. Gupta, P. Lenka, A. Ekbal, P. Bhattacharyya, A unified framework for multilingual and code-mixed visual question answering, in: K.-F. Wong, K. Knight, H. Wu (Eds.), *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, Association for Computational Linguistics, Suzhou, China, 2020, pp. 900–913. URL: <https://aclanthology.org/2020.acl-main.90>.
- [15] H. Chase, Langchain, ??? URL: <https://github.com/langchain-ai/langchain>.
- [16] M. development contributors, L. Beyer*, A. Steiner*, A. S. Pinto*, A. Kolesnikov*, X. Wang*, X. Zhai*, D. Salz, M. Neumann, I. Alabdulmohsin, et al., *Paligemma* (2024). URL: <https://www.kaggle.com/m/23393>. doi:10.34740/KAGGLE/M/23393.
- [17] P. Esser, S. Kulal, A. Blattmann, R. Entezari, J. Müller, H. Saini, Y. Levi, D. Lorenz, A. Sauer, F. Boesel, D. Podell, T. Dockhorn, Z. English, K. Lacey, A. Goodwin, Y. Marek, R. Rombach, Scaling rectified flow transformers for high-resolution image synthesis (2024).
- [18] K. Pipatanakul, P. Jirabovonvisut, P. Manakul, S. Sripaisarnmongkol, R. Patomwong, P. Chokchainant, K. Tharnpipitchai, Typhoon: Thai large language models, 2023. URL: <https://arxiv.org/abs/2312.13951>. arXiv:2312.13951.

- [19] AI@Meta, Llama 3 model card (2024). URL: https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md.
- [20] T. Dettmers, A. Pagnoni, A. Holtzman, L. Zettlemoyer, Qlora: Efficient finetuning of quantized llms, 2023. URL: <https://arxiv.org/abs/2305.14314>. arXiv:2305.14314.
- [21] P. Pengpun, C. Udomcharoenchaikit, W. Buaphet, P. Limkonchotiwat, Seed-free synthetic data generation framework for instruction-tuning LLMs: A case study in Thai, in: X. Fu, E. Fleisig (Eds.), Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 438–457. URL: <https://aclanthology.org/2024.acl-srw.38>. doi:10.18653/v1/2024.acl-srw.38.
- [22] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, Microsoft coco: Common objects in context, 2015. URL: <https://arxiv.org/abs/1405.0312>. arXiv:1405.0312.
- [23] Theerasit, Ai cooking image captioning, 2024. URL: <https://kaggle.com/competitions/ai-cooking-image-captioning>.
- [24] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, Robust speech recognition via large-scale weak supervision, 2022. URL: <https://arxiv.org/abs/2212.04356>. arXiv:2212.04356.
- [25] K. Thangthai, S. Thatphithakkul, V. Chunwijitra, Ai-cooking asr dialects, Kaggle (2024). URL: <https://www.kaggle.com/competitions/ai-cooking-asr-dialects>.