Cooperative Education Report

Comparative Analysis of Mask-Wearing Stance and Emotional Intensity in Thai and USA YouTube Comments during the COVID-19 Pandemic

At Nara Institute of Science and Technology

Mr. Supakrit Aphonmaeklong
Student ID 6410545789

Academic Year 2024

Department of Computer Engineering, Faculty of Engineering
Kasetsart University

Abstract

This internship report is part of the curriculum for the Department of Software and Knowledge Engineering at Kasetsart University's Faculty of Engineering. During my internship from May 13 to August 9, 2024, I had the opportunity to conduct research at the Social Computing Lab at Nara Institute of Science and Technology under the supervision of Professor Eiji Aramaki. My research project, titled "Comparative Analysis of Mask-Wearing Stance and Emotional Intensity in Thai and USA YouTube Comments during the COVID-19 Pandemic," involved collecting and analyzing YouTube comments to determine public sentiment and emotional intensity towards mask-wearing. I utilized advanced natural language processing techniques and fine-tuned state-of-the-art models such as GEMMA 2 and EmoLLama-chat-7b for stance detection and emotion analysis. Significant cultural differences and media influence on public opinion between the two countries were revealed. This experience provided me with valuable skills in data collection, model fine-tuning, sentiment analysis, and cross-cultural research methodologies, significantly contributing to my academic and professional development.

Acknowledgement

I would like to express my gratitude to Nara Institute of Science and Technology for providing excellent research facilities, including the lab and Ubuntu GPU servers, which enabled the fine-tuning of large language models. My sincere thanks go to Professor Eiji Aramaki for his constructive ideas, guidance, and support throughout this research project. I am also thankful to the members of the Social Computing Lab, including master's and PhD students, for their support, advice, and willingness to consult with me. Special thanks to Kiki Ferawati, a PhD student in the lab, for providing her crowd-sourced dataset on mask-wearing Twitter stance, which was essential for training my model.

Mr. Supakrit Aphonmaeklong
Reporter

The last date of co-op
09 / August / 2024

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Motivations and Importance

The motivation behind this research topic stems from the global impact of the COVID-19 pandemic, which has highlighted the significance of public health measures such as mask-wearing. Understanding public sentiment and emotional responses to these measures is crucial for developing effective communication strategies and policies. By analyzing YouTube comments from viewers in Thailand and the USA, this study aims to uncover cultural and emotional differences in attitudes towards mask-wearing. The insights gained from this research are important for policymakers and health communicators to tailor their messages to specific audiences, ultimately improving public compliance with health directives and mitigating the spread of the virus.

1.2 Objectives

1. **Stance Detection:** To determine the public stance towards mask-wearing from YouTube comments in Thailand and the USA using advanced NLP models.
2. **Emotion Analysis:** To analyze the emotional intensity (fear, anger, joy) associated with mask-wearing comments and how these emotions vary between the two countries.
3. **Temporal Trends:** To examine how public sentiment and emotional intensity towards mask-wearing have evolved from 2020 to 2022.
4. **Media Influence:** To compare the stance and emotional intensity of comments on videos from different news publishers within each country.
5. **Correlation Analysis:** To explore the relationship between public sentiment, emotional intensity, and daily COVID-19 case counts in Thailand and the USA.

1.3 Scope of Work

My research focuses on understanding public sentiment and emotional intensity towards mask-wearing during the COVID-19 pandemic. YouTube comments from major news publishers in Thailand and the USA will be collected using the YouTube API. The GEMMA 2 model will be fine-tuned for stance detection using the SMM4H 2022 Task 2 dataset, and the EmoLLama-chat-7b model will be used for emotion analysis, focusing on fear, anger, and joy, with Thai comments translated into English. Statistical analysis, including OLS regression, will assess the impact of emotional intensity and stance on daily COVID-19 cases and deaths in both countries. Comparative analysis will evaluate public sentiment and emotional responses between Thailand and the USA, as well as the influence of different news publishers. The findings will be presented through detailed tables and visualizations, providing insights for policymakers and public health communicators.

1.4 History and Detail of Company

- Name and Location
  - Nara Institute of Science and Technology
  - 630-0192 Nara, Ikoma, Takayamacho, 8916-5
- Company Profile
  - Nara Institute of Science and Technology (NAIST) is a Japanese national university situated in Kansai Science City, which lies at the intersection of Nara, Osaka, and Kyoto. Established in 1991, NAIST originally comprised graduate schools focused on information science, biological sciences, and materials science. In 2018, the university reorganized to maintain its research in these fields while fostering interdisciplinary research and education. Now operating as a single graduate school, NAIST is dedicated to advancing frontier research and preparing students to become future leaders in science and technology.
- Organization and management model
  - The Nara Institute of Science and Technology (NAIST) organizes its academic structure into three graduate schools, each encompassing a variety of specialized departments and research centers. The Graduate School of Information Science includes the Departments of Computational Neuroscience, Cyber Resilience, Data Science and Knowledge Engineering, Human-Information Interaction, Mathematical Informatics, Robotics, and Software Engineering. The Graduate School of Biological Sciences comprises the Departments of Biomedical Science, Plant Developmental Biology,

Systems Biology, Structural Biology, Infection Biology, Neuroscience, and Cell Biology. The Graduate School of Materials Science features the Departments of Applied Physics, Materials Design Innovation Engineering, Materials Chemistry, Molecular Devices, and Quantum Materials. This structure promotes interdisciplinary research and education, encouraging collaboration across departments to drive innovation and tackle complex scientific challenges.

  - For my internship, I worked under the supervision of Professor Eiji Aramaki from the Social Computing Lab at the Nara Institute of Science and Technology (NAIST).

- My Position
  - A Research Internship student at the Social Computing Lab, NAIST, working on sentiment analysis of YouTube comments regarding mask-wearing.
- Supervisor Position
  - Eiji Aramaki, Ph. D, professor at NAIST (Nara Institute of Science and Technology)
- Internship Period
  - 13 May 2024 to 9 August 2024

1.5 Expected Benefits

During my internship, I benefited in several key areas. I learned to use open-source Hugging Face models [5] and fine-tuned large language models (LLMs) with custom tasks and datasets. I gained practical knowledge of basic procedures for NLP tasks, as well as effective visualization and analysis techniques using Python libraries. I also learned to work with remote servers to train large machine learning models and collected and fetched YouTube comment data on thousands of videos related to COVID-19. Collaborating with Japanese colleagues provided an opportunity to learn basic Japanese language. These experiences have collectively enriched my technical capabilities and cultural understanding, preparing me for future projects in the field of artificial intelligence and beyond.

Chapter 2
Background Knowledge and Related Work

2.1 Background Knowledge

2.1.1 Machine Learning

Machine learning is a branch of artificial intelligence that involves training algorithms to learn from and make predictions on data. It is the backbone of many modern technologies, including NLP, image recognition, and autonomous systems. The basic steps in a machine learning process include data collection, data preprocessing, model selection, training, evaluation, tuning, and prediction. Data collection involves gathering relevant data from various sources. Data preprocessing includes cleaning and preparing data for analysis by handling missing values, normalization, and transforming data into a suitable format. Model selection involves choosing an appropriate machine learning algorithm based on the problem and data characteristics. During training, the data is fed into the model, allowing it to learn patterns and relationships. Evaluation assesses the model's performance using metrics like accuracy, precision, recall, and F1 score. Tuning adjusts the model's parameters to enhance performance. Finally, prediction uses the trained model to make predictions on new, unseen data.

2.1.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a subfield of artificial intelligence that focuses on the interaction between computers and human languages, enabling machines to understand, interpret, and generate human language. Modern NLP techniques include tokenization, part-of-speech tagging, named entity recognition (NER), sentiment analysis, and language modeling. Transformative advancements in NLP have been driven by transformer models like BERT, GPT, and RoBERTa, which utilize attention mechanisms and transfer learning to enhance language processing accuracy and efficiency. These techniques are crucial for applications such as language translation, speech recognition, and text summarization.

2.1.3 Large Language Models (LLMs)

LLMs are powerful machine learning models trained on vast amounts of text data. They can perform a variety of NLP tasks with high accuracy. In this project, models like GEMMA 2 and EmoLLama-chat-7b are fine-tuned to analyze public sentiment and emotional intensity in YouTube comments. LLMs are capable of understanding and generating human-like text, making them ideal for complex NLP tasks.

2.1.4 PyTorch

PyTorch is an open-source machine learning library widely used for developing and training deep learning models [8]. It provides flexible and efficient tools for implementing complex neural networks and is integral to this project for model fine-tuning and experimentation. PyTorch’s dynamic computation graph and extensive support for GPU acceleration make it ideal for handling large-scale NLP tasks.

2.2 Related Work

Prior to this internship, I have been using Python as my main programming language since my first year. I have developed a solid foundation in Python programming through courses such as Computer Programming I and Computer Programming II. Additionally, in my third year, I gained a deeper understanding of data analytics in my Data Analytics class, where I learned about machine learning methodologies and classical machine learning models like logistic regression, decision trees, and model validation techniques. I also attended a Deep Learning class (01219494), which provided me with knowledge about the basics of neural networks and typical deep learning concepts such as backpropagation, convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory (LSTM) networks, and transformer model architectures. This foundational knowledge has been crucial in preparing me for future work with state-of-the-art models like BERT, LLaMA, and GEMMA.

Moreover, I have read "Emotion Analysis of Writers and Readers of Japanese Tweets on Vaccinations" [9], which provided me with a foundational understanding of data collection and sentiment analysis methodologies. The study's techniques in fine-tuning a BERT model for emotion intensity prediction and its subsequent application to a large dataset of Japanese tweets informed my approach to using advanced NLP models.

Chapter 3
Methodology

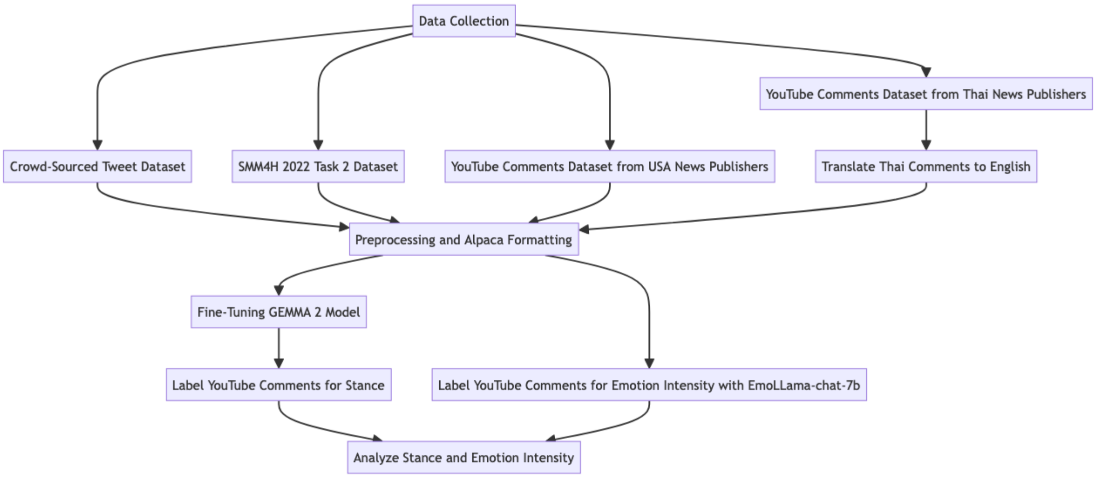


Figure 1: High-level Research Methodology

This section details the comprehensive methodology employed in our study to analyze stance and emotional intensity towards mask-wearing during the COVID-19 pandemic. Figure 1 illustrates the high-level methodology of our research project. Below, we present and explain the key processes in our methodology.

3.1 Data Collection

To investigate the stance and emotional intensity towards mask-wearing during the COVID-19 pandemic, we utilized four distinct datasets: two for training and validation, and two for prediction and analysis of YouTube comments from Thailand and the USA.

Claim/Topic	Stance		
	AGAINST	FAVOR	NONE
Crowd-Sourced Tweet Dataset			
face masks	201	506	393
the SMM4H 2022 Task 2: train set			
face masks	324	652	343
school closures	217	526	307
stay at home orders	333	168	686
the SMM4H 2022 Task 2: test set			
face masks	48	190	235
school closures	23	205	226
stay at home orders	116	92	352

Table 1: Summary of Stance Counts in the Crowd-Sourced Tweet Dataset and smm4h 2022 Task 2 train and test sets.

3.1.1 Crowd-Sourced Tweet Dataset

The first dataset is a crowd-sourced collection of 1,100 tweets about mask-wearing, provided by a laboratory member from the Social Computing Lab. This dataset includes stance labels categorized for ternary classification (0: against, 1: not against, ?/blank: unclear). In cases of annotator disagreement on stance, tweets were marked as "unclear" (?). Its stance distribution is shown in Table 1. This dataset was used exclusively for training the model.

3.1.2 SMM4H 2022 Task 2 Dataset

The second dataset is derived from the SMM4H 2022 Task 2, which focuses on stance detection and premise classification in COVID-19-related tweets [2]. This dataset consists of 5,550 annotated tweets related to three claims about COVID-19 mandates: face masks, school closures, and stay-at-home orders. Table 1 displays stance distribution from 3 topics on the train and test sets. For this study, we utilized the training set (3,556 tweets) to fine-tune the GEMMA 2 model, and its test set to validate the model's performance.

Dataset	Total Comments	Comments per Publisher
Thailand		
TNN	1539	513
Rueng Lao Chao Nee		513
Thairath TV Originals		513
USA		
ABC News	9228	3046
CNN		3046
Fox News		3046

Table 2: Summary of YouTube comments dataset distribution per news publisher in Thailand and the USA.

Dataset	Comments per Year		
	2020	2021	2022
Thailand	387	818	334
USA	3308	3678	2152

Table 3: Summary of YouTube comments distribution per year for Thailand and the USA.

3.1.3 YouTube Comments Dataset from USA News Publishers

The third dataset comprises YouTube comments collected from major news publishers in the USA, including ABC News, CNN, and Fox News, using the YouTube API [4]. Comments were gathered from COVID-19-related videos between 2020 and 2022, specifically selecting those containing the keywords “mask” and “wear.” Basic text processing, such as URL removal, ensured the relevance of these keywords. Each publisher’s dataset was balanced with 3,046 comments, as shown in Table 2. All comments sum up to 9228. The dataset has 3308 comments in 2020, 3678 in 2021, and 2152 in 2022 as shown in Table 3.

3.1.4 YouTube Comments Dataset from Thai News Publishers

The fourth dataset consists of YouTube comments collected from major news publishers in Thailand, including TNN, Rueng Lao Chao Nee, and Thairath TV, using the YouTube API [4]. Comments were gathered from COVID-19-related videos between 2020 and 2022, specifically selecting those containing the Thai keywords for “mask” (หน้ากาก, มาส์ก, แมส, มาส์ค, แมสก์) and “wear” (ใส่). Basic text processing, such as URL removal, ensured the relevance of these keywords. To analyze these comments, we utilized the Gemini 1.5 pro API from Google Cloud Platform to translate the Thai YouTube comments into English text [3]. Each publisher’s dataset was balanced with 513 comments, as shown in Table 2. All comments sum up to 1539. The dataset has 387 comments in 2020, 818 in 2021, and 334 in 2022 as shown in Table 3.

3.2 Preprocessing and Alpaca Formatting

After collecting the datasets, the next step involved preprocessing and formatting the data for model training. Each dataset was cleaned and structured following the Alpaca format, ensuring consistency and clarity in model training. Each training example was structured as follows:

Instruction:
[Instruction on what to do with the input]
Input:
[The input text or data]
Output:

[The expected output or response]
This format included explicit instructions, relevant inputs (tweets or comments), and expected stance outputs, thereby streamlining the training process and improving model interpretation accuracy.

3.3 Fine-Tuning GEMMA 2 for Stance Analysis

To adapt the GEMMA 2 9b model for stance detection in mask-wearing comments during the COVID-19 pandemic, we employed a parameter-efficient fine-tuning approach leveraging 4-bit quantization. This model was obtained from Unsloth [10]. This aimed to optimize computational resources while enhancing the model's ability to discern stances in social media comments.

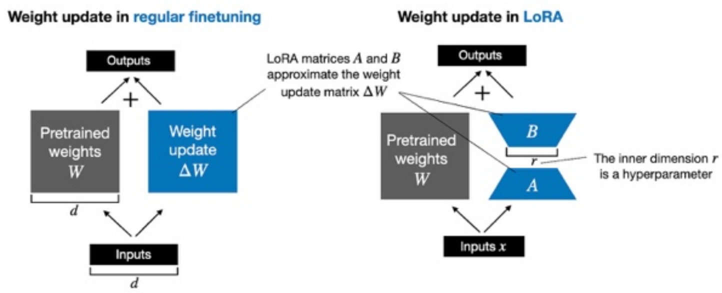


Figure 2: LoRA weights update

We utilized Low-Rank Adaptation (LoRA) to fine-tune specific components of the model architecture, such as the query, key, value, and output projections within the transformer layers [6]. LoRA decomposes the weight update matrix into two smaller matrices, significantly reducing the number of trainable parameters. As shown in Figure 2, instead of updating the full weight matrix W, LoRA updates two smaller matrices A and B such that W = W_0 + \alpha \cdot AB, where W_0 is the original weight matrix, and \alpha is a scaling factor. This allows for efficient training with lower computational and memory requirements. To mitigate overfitting, a dropout rate of 0.4 was incorporated, and the rank of adaptation layers was set to 32, balancing model complexity and training efficiency. This approach ensures faster convergence and reduced training time while maintaining the model's performance and generalizability.

Training data was organized in the Alpaca format for consistency and clarity. Each example included instructions, relevant inputs (tweets or comments), and expected stance outputs. The training dataset included crowd-sourced tweets and the SMM4H Task 2 training set. To address class imbalance, the dataset was balanced across the three stance categories—Favor, Against, and Neutral—by under sampling the majority classes.

The fine-tuning process used the SFTTrainer framework from PyTorch [8] with configurations to optimize performance. Training was limited to one epoch due to the large dataset size, with a batch size of two for training and four for evaluation. A learning rate of 1×10^{-4} ensured stable training. The AdamW optimizer with 8-bit precision enhanced memory efficiency. Model performance was assessed every 100 steps, with the best model saved based on evaluation loss. Training was conducted on an NVIDIA A100-PCIE-40GB GPU, leveraging its memory capacity for efficient and stable training.

3.4 Labeling YouTube Comments for Stance

Once the GEMMA 2 model was fine-tuned, it was employed to label the YouTube comment datasets from Thailand and the USA. Each comment was categorized into one of the three stance categories—Favor, Against, or Neutral—based on the model's predictions. This step was crucial for the subsequent analysis of stance distribution and emotional intensity in the YouTube comments.

3.5 Labeling YouTube Comments for Emotion Intensity with EmoLLama-chat-7b

For the task of analyzing the emotional intensity of YouTube comments, specifically focusing on joy, anger, and fear, we utilized the EmoLLama-chat-7b model, part of the EmoLLMs project [7]. This model, finetuned based on the LLaMA2-chat-7B foundation model, provides comprehensive affective analysis capabilities. The model was employed to assign numerical values representing the intensity of emotions expressed in the comments.

3.6 Analyze stance and emotion intensity

Finally, we evaluated mask-wearing stance and emotional intensity, and the influence of news publishers on mask-wearing by analyzing the labeled YouTube comments. Moreover, we also conducted statistical correlation analysis on the stance and emotion intensity towards COVID-19 daily cases and deaths. The full analysis results will be shown in Chapter 4, under the Results subsection.

Chapter 4

Results and Analysis

4.1 Results

4.1.1 Fine-tuning Results

4.1.1.1 Performance of GEMMA 2 LLM

The GEMMA 2 9b model with 4-bit quantization was fine-tuned to enhance its stance classification capabilities, specifically focusing on the mask-wearing debate during the COVID-19 pandemic. This model was evaluated using the SMM4H Task 2 dataset test set, which is renowned for its challenging stance detection scenarios. The primary metric for evaluating the fine-tuned model was the macro F1 score, which balances the precision and recall across all classes. The GEMMA 2 LLM achieved an overall macro F1 score of 0.73, indicating a robust ability to correctly identify and classify stances from text data. Table 4 presents the detailed classification report for GEMMA 2 LLM on the SMM4H Task 2 test set, including precision, recall, and F1 scores for each stance category.

Stance	Precision	Recall	F1-score	Support
Against	0.65	0.59	0.62	187
Favorable	0.76	0.77	0.77	487
Neutral	0.80	0.81	0.81	813
Accuracy	0.77			
Macro avg	0.74	0.73	0.73	1487
Weighted avg	0.77	0.77	0.77	1487

Table 4: Overall classification report for GEMMA 2 LLM on the SMM4H Task 2 test set.

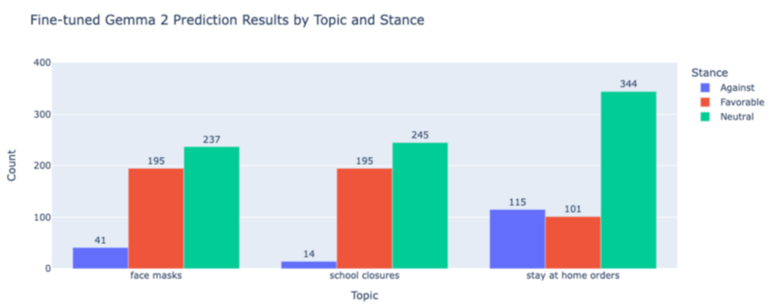


Figure 3: Fine-tuned GEMMA 2 prediction results by topic and stance.

To provide a visual representation of the GEMMA 2 LLM's classification performance, Figure 3 illustrates the distribution of the model's predictions by topic and stance. This figure highlights the model's prediction counts for each stance category across different topics.

Figure 4 illustrates the confusion matrix for the fine-tuned GEMMA 2 LLM on the SMM4H Task 2 test set. The matrix shows the model's performance in classifying stances as Against, Favor, or Neutral. Correct classifications appear on the diagonal: 111 Against, 375 Favor, and 642 Neutral. Misclassifications are shown off-diagonal, such as 13 Against instances predicted as Favor and 323 Neutral instances predicted as Favor. The analysis indicates the model's bias towards predicting Favor, as many Neutral instances are misclassified as Favor. Additionally, the model struggles to

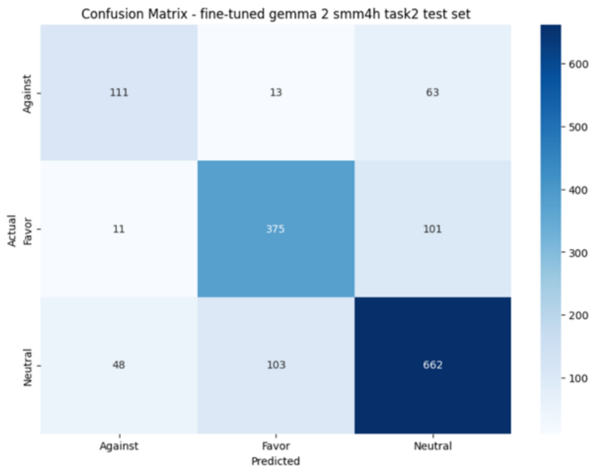


Figure 4: Confusion Matrix for GEMMA 2 LLM on the SMM4H Task 2 test set.

Topic	Precision	Recall	F1-score
School Closures	0.56	0.55	0.55
Stay at Home Orders	0.72	0.74	0.73
Face Masks	0.76	0.74	0.75

Table 5: Macro F1 scores for each topic in the SMM4H Task 2 test set.

distinguish Against stances, with several being misclassified as Neutral or Favor. These issues highlight the need for better differentiation features and potential dataset rebalancing.

Table 5 summarizes the macro F1 scores for each topic, providing a detailed view of the model's capabilities in handling different aspects of the COVID-19 related topics. The "Stay at Home Orders" and "Face Masks" topics have high macro F1 scores of 0.73 and 0.75, respectively, while "School Closures" have a lower macro F1 score of 0.55.

4.1.1.2 Comparison with Other Model

Model	Macro F1 Score
Random	0.268
General-domain BERT	0.464
COVID-Twitter-BERT	0.601
BART + Syntax Features	0.45
DAN-BERT	0.581
RoBERTa Baseline	0.450
GEMMA 2	0.73

Table 6: Comparison of macro F1 scores between GEMMA 2 and other models on the SMM4H Task 2 test set.

Claim/Topic	RoBERTa Baseline F1 Score (Test Set)
Face Masks	0.439
Close School	0.345
Home Orders	0.566

Table 7: RoBERTa baseline model macro F1 scores on the SMM4H Task 2 test set.

To contextualize the performance of our GEMMA 2 fine-tuned model, we compared its stance classification metrics against various models that have classification metrics on the SMM4H Task 2 test set. Table 6 summarizes the macro F1 scores for stance classification from different models.

The GEMMA 2 model, fine-tuned for stance detection, achieved a macro F1 score of 0.73, surpassing the performance of other models on the SMM4H Task 2 test set as shown in Table 6. There are five models mentioned in [2], consisting of Random, General-domain BERT, COVID-Twitter-BERT, BART + Syntax Features, and DAN-BERT. The random baseline model, which assigns labels randomly without any learned understanding, achieved a macro F1 score of 0.268. The BERT-base model, a general-domain pre-trained model consisting of 12 layers of Transformer encoder with 12 attention heads and 110M parameters, achieved a macro F1 score of 0.464. Although BERT's architecture allows it to produce robust context-based representations, its performance on this specialized task remains moderate without domain-specific fine-tuning. COVID-Twitter-BERT, a BERT-large-uncased model pre-trained on a large corpus of tweets specifically related to COVID-19, achieved a macro F1 score of 0.601. This domain-specific pre-training significantly enhances its ability to understand and classify COVID-19 related stances accurately. The BART model, enhanced with syntactic features and based on an encoder-decoder transformer architecture, resulted in a macro F1 score of 0.45. Although BART's pre-training task helps it handle noisy and ambiguous text, its performance lags models specifically pre-trained on COVID-19 data. DAN-BERT, which incorporates a dual-view architecture to learn both subjective and objective features of texts, achieved a macro F1 score of 0.581. This model's ability to separate different types of textual information contributes to its strong performance in stance detection tasks. The RoBERTa model from [1], a robustly optimized variant of BERT, achieved a test F1 score of 0.450 (Table 7). Despite RoBERTa's improved training strategies, it still falls short compared to more specialized models. Additionally, the GEMMA 2 model demonstrated a better F1 score on the face masks topic compared to the RoBERTa baseline, further highlighting its reliable performance on the face mask topic, which is the focus of this research paper. This comparison underscores the efficacy of the GEMMA 2 model in handling stance detection tasks and emphasizes its potential as a valuable tool for public sentiment analysis, especially in capturing nuanced information from social media text.

4.1.2 YouTube Comments Mask Wearing Stance Analysis

4.1.2.1 Overall Stance Analysis

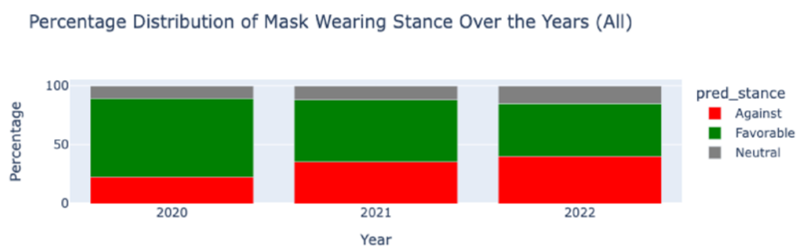
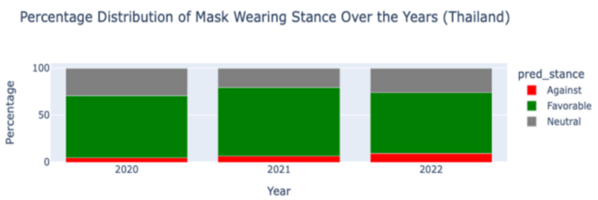


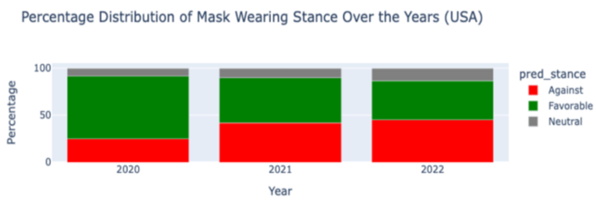
Figure 5: Percentage distribution of mask wearing stance over the years.

The analysis of the stance on mask-wearing over the years, presented in Figure 5, reveals distinct trends in public opinion from 2020 to 2022. During this period, the favorable stance towards mask-wearing consistently dominated, representing more than half of the overall stance distribution each year. However, there is a noticeable gradual decline in the proportion of favorable stances annually, while the percentages of against and neutral stances have slightly increased each year. These trends reflect the evolving public attitudes towards mask-wearing, indicating a sustained majority support with minor annual shifts in opposition and neutrality.

4.1.2.2 Thailand vs. USA



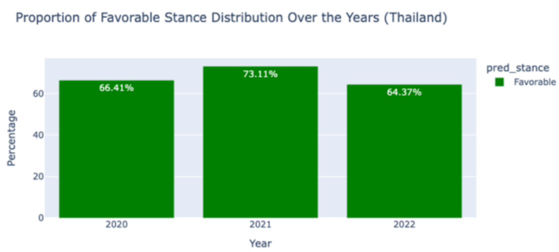
(a) Thailand Stance Yearly Distribution



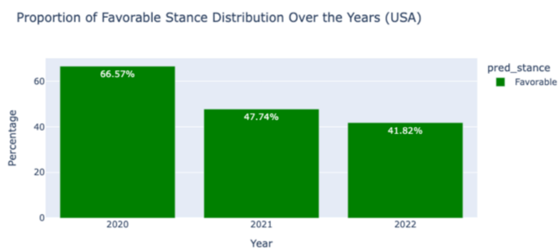
(b) USA Stance Yearly Distribution

Figure 6: Percentage distribution of mask wearing stance over the years in Thailand and the USA.

Figure 6 presents the yearly distribution of mask-wearing stances in both Thailand and the USA from 2020 to 2022. In Thailand, the favorable stance significantly dominated other stances each year. In 2022, the favorable stance was over six times greater than the against stance, indicating that Thai people generally had a positive view on wearing masks. Conversely, in the USA, the favorable stance declined from 66.57% in 2020 to 41.82% in 2022, indicating a reduction in public support for mask-wearing over time. This contrast shows a consistent support in Thailand compared to a diminishing support in the USA.



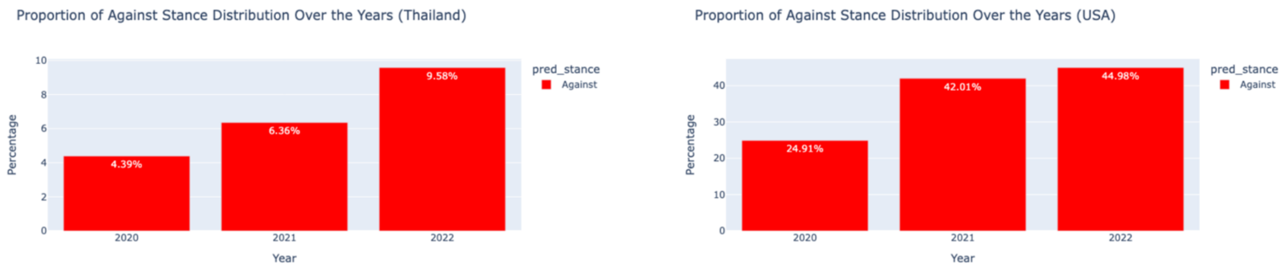
(a) Thailand Favorable Stance Yearly Distribution



(b) USA Favorable Stance Yearly Distribution

Figure 7: Proportion of Favorable stance distribution over the years in Thailand and the USA.

Figure 7 shows the proportion of favorable stance distribution over the years in Thailand and the USA. In Thailand, the favorable stance peaked at 73.11% in 2021, aligning with the peak of COVID-19 daily cases in Thailand during that year, suggesting an increase in public support for mask-wearing as the pandemic intensified. In the USA, the percentage of favorable comments decreased from 66.57% in 2020 to 47.74% in 2021, and further to 41.82% in 2022, reflecting a diminishing support for mask-wearing. This comparison highlights the contrasting trends in public support for mask-wearing between the two countries.



(a) Thailand Against Stance Yearly Distribution

(b) USA Against Stance Yearly Distribution

Figure 8: Proportion of Against stance distribution over the years in Thailand and the USA.

Figure 8 illustrates the proportion of against stance distribution over the years in Thailand and the USA. In Thailand, although the against stance increased each year, it remained below 10% even at its highest in 2022. This relatively low opposition further underscores the general acceptance and support for mask-wearing among the Thai population. In the USA, the against stance increased from 24.91% in 2020 to 42.01% in 2021, and further to 44.98% in 2022, indicating a growing opposition to mask mandates. This contrast reveals a significant difference in the level of opposition to mask-wearing between the two countries.

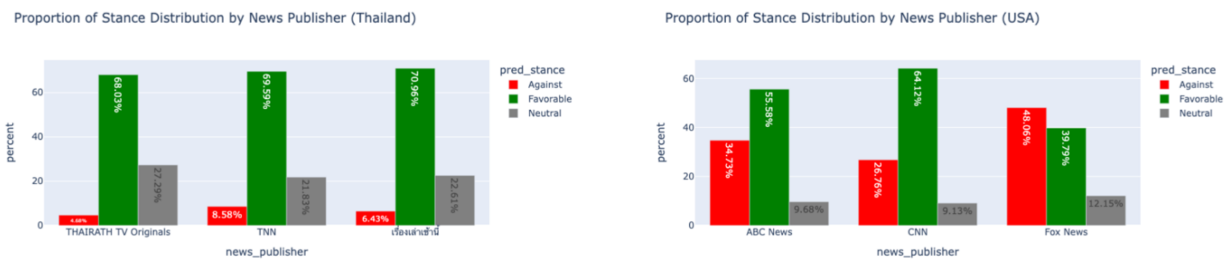


Figure 9: Proportion of stance distribution by news publishers in Thailand and the USA.

Figure 9 highlights the stance distribution by news publishers in Thailand and the USA. In Thailand, the stance distribution by news publishers is more consistent, with high proportions of favorable stances across all major news publishers. In contrast, the USA shows a notable difference between CNN and Fox News. CNN, which is often perceived as having a more liberal audience, shows a higher proportion of favorable stance towards mask-wearing at 64.12%, compared to Fox News, which is perceived as having a more conservative audience, with a favorable stance of only 39.79%. Conversely, the proportion of against stance is significantly higher for Fox News at 48.06%, compared to CNN at 26.76%. This distribution is notably different from that in Thailand, where public opinion on mask-wearing is more uniformly positive across different news publishers.

4.1.3 YouTube Comments Emotional Intensity Analysis

4.1.3.1 Overall Emotional Intensity Analysis

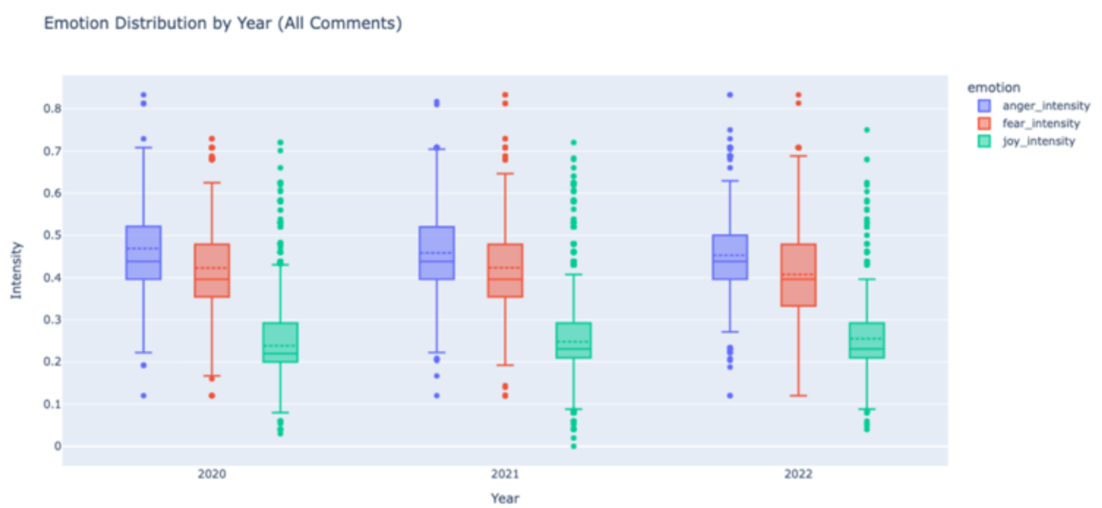
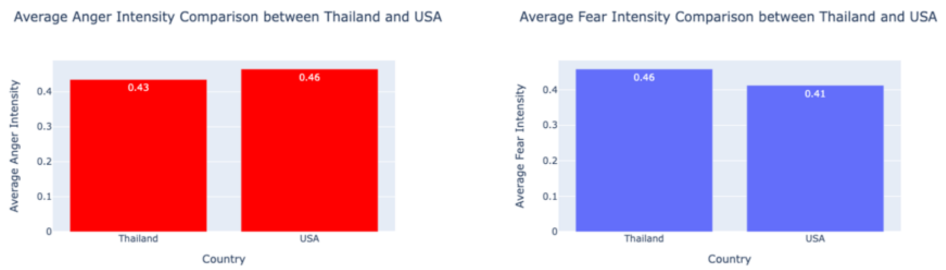


Figure 10: Emotion intensity distribution by year.

The analysis of emotional intensity in comments regarding mask-wearing during the COVID-19 pandemic reveals distinct trends over the years. As depicted in Figure 10, the emotional intensity shows a stable pattern where anger is consistently higher than fear, and fear is higher than joy. There is a slight, consistent rise in joy from approximately 0.23 in 2020 to 0.25 by 2022. This indicates a subtle shift towards more positive sentiments over time, despite the ongoing pandemic

4.1.3.2 Thailand vs. USA

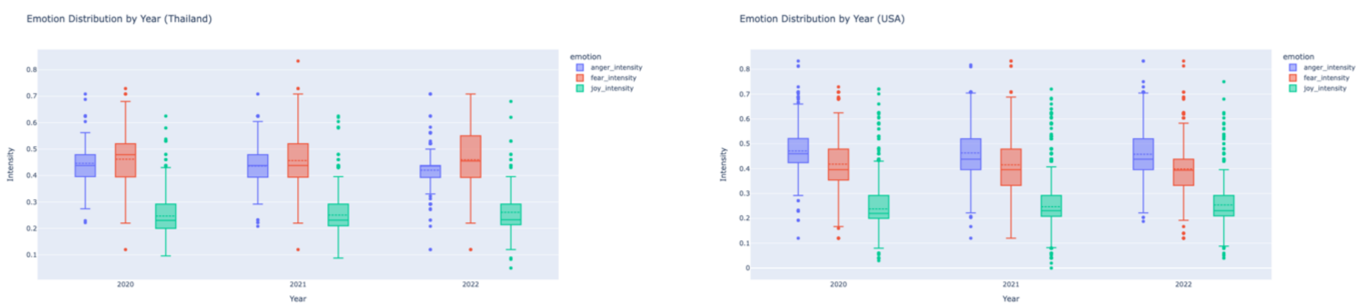
4.3.1 Thailand vs. USA



(a) Anger Intensity Comparison between Thailand and USA (b) Fear Intensity Comparison between Thailand and USA

Figure 11: Comparison of anger and fear intensity between Thailand and the USA.

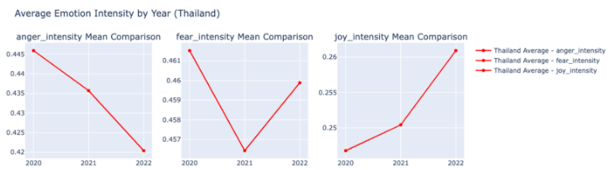
The analysis of emotional intensity in comments regarding mask-wearing during the COVID-19 pandemic unveils distinct emotional responses between Thailand and the USA. Figure 11a presents the comparison of anger intensity between Thailand and the USA. The USA exhibits slightly higher anger intensity at 0.46 compared to Thailand's 0.43. This indicates that the public in the USA has expressed more anger in comments regarding mask-wearing than in Thailand. Figure 11b illustrates the comparison of fear intensity between Thailand and the USA. Thailand demonstrates a significantly higher fear intensity at 0.46 compared to the USA's 0.41. This suggests that the Thai public has expressed a greater level of concern and anxiety towards the pandemic.



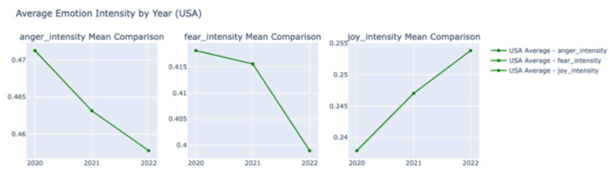
(a) Box Plot Emotion Intensity by Year (Thailand) (b) Box Plot Emotion Intensity by Year (USA)

Figure 12: Box plot emotion intensity by year in (a) Thailand and (b) USA.

Figure 12 provides a comprehensive overview of emotion intensity through box plots, highlighting the variability and median trends in anger, fear, and joy intensities in Thailand from 2020 to 2022. Fear consistently emerges as the most intense emotion each year, followed by anger and joy. In contrast, the box plot for the USA shows that anger is the most pronounced emotion each year, followed by fear and joy. This indicates differing emotional landscapes between the two countries, with Thai comments showing higher fear intensity and US comments showing higher anger intensity.



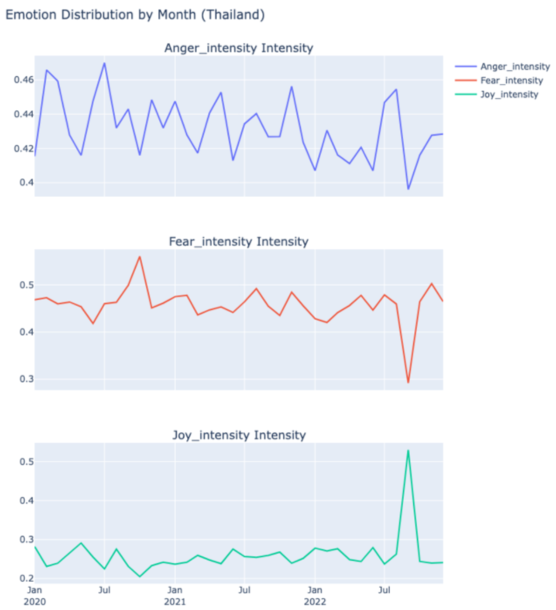
(a) Average Emotion Intensity by Year (Thailand)



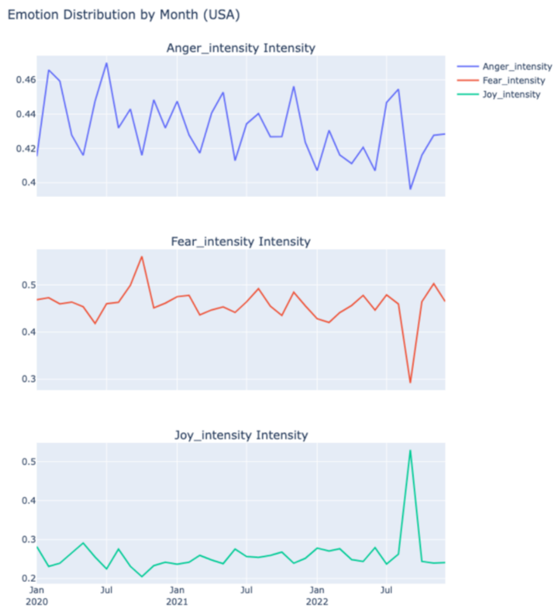
(b) Average Emotion Intensity by Year (USA)

Figure 13: Average emotion intensity by year in (a) Thailand and (b) USA.

Figure 13 details the annual changes in emotions in Thailand (a) and the USA (b). In Thailand, there is a gradual decrease in anger intensity year-over-year, showing a downward trend from 0.4459 in 2020 to 0.4204 in 2022. Joy intensity shows a slight increase annually, starting at 0.2467 in 2020, rising to 0.2504 in 2021, and reaching 0.2609 in 2022. Fear intensity initially peaked at 0.4615 in 2020, decreased slightly to 0.4560 in 2021, and rose again to 0.4598 in 2022. In contrast, the annual changes in the USA show that anger intensity decreases from 0.4712 in 2020 to 0.4577 in 2022. Joy intensity shows a mild upward trend, increasing from 0.2370 in 2020 to 0.2530 in 2022. Fear intensity decreases notably from 0.4156 in 2021 to 0.3980 in 2022.



(a) Average Emotion Intensity by Month (Thailand)



(b) Average Emotion Intensity by Month (USA)

Figure 14: Average Emotion Intensity of Anger, Fear, and Joy by Month in (a) Thailand and (b) USA.

Figure 14a provides insights into the monthly variations in emotion intensities in Thailand. Significant peaks in fear intensity occurred in September and October 2020, driven by concerns over a potential second wave and inconsistent mask usage among the population, including foreigners and migrants. In July and August 2022, a notable peak in anger intensity occurred due to public dissatisfaction with the authorities' decision to relax mask mandates and perceived discrepancies in mask-wearing compliance between tourists and locals. Similarly, Figure 14b shows the monthly variations in the USA, illustrating how public emotions react in real-time to evolving pandemic conditions and governmental responses.

4.1.4 Correlation Analysis between Stance, Emotional Intensity, and COVID-19 Cases.

4.1.4.1 Thailand

Variable	Coefficient (Cases)	P-value (Cases)	Coefficient (Deaths)	P-value (Deaths)
Constant	4044.1874	0.020	26.1741	0.035
Fear Intensity	2415.7588	0.236	24.1949	0.095
Joy Intensity	5099.7672	0.047	19.4465	0.286
Anger Intensity	-4267.2184	0.115	-27.0450	0.160

Table 8: OLS regression results for new covid-19 cases and deaths in Thailand.

For Thailand, we merged the stance and emotion data with daily COVID-19 case counts obtained from the WHO dataset [11]. The merged dataset was analyzed using Ordinary Least Squares (OLS) regression to assess the impact of emotional intensity and stance on the number of new COVID-19 cases and deaths. Table 8 shows the OLS regression results for new cases and deaths in Thailand. The results indicate that joy intensity has a statistically significant positive impact on the number of new COVID-19 cases, with a coefficient of 5099.7672 (p = 0.047). This suggests that higher joy intensity in public sentiment may correlate with a perceived reduction in threat, potentially leading to more relaxed behaviors and increased case counts. Fear and anger intensities, along with stance, did not show significant effects on new cases or deaths.

4.1.4.2 USA

Variable	Coefficient (Cases)	P-value (Cases)	Coefficient (Deaths)	P-value (Deaths)
Constant	81940.000	0.000	954.520	0.000
Fear Intensity	21450.000	0.313	-248.242	0.019
Joy Intensity	43910.000	0.063	-18.346	0.876
Anger Intensity	-9354.000	0.693	430.870	0.000
Pred Stance	-8163.673	0.000	15.183	0.098

Table 9: OLS regression results for new covid-19 cases and deaths in the USA.

Similarly, for the USA, we merged the stance and emotion data with daily COVID-19 case counts and performed OLS regression analysis. Table 9 shows the OLS regression results for new cases and deaths in the USA. The results indicate that pred_stance significantly affects the number of new COVID-19 cases. The coefficient for pred_stance is -8163.673, meaning a favorable stance on mask-wearing is associated with a decrease of approximately 8163 new cases compared to a neutral stance. This effect is highly significant (p = 0.000). Fear intensity and anger intensity significantly affect the number of new COVID-19 deaths. The coefficient for fear intensity is -248.242, meaning higher fear intensity is associated with a decrease of approximately 248 new deaths (p = 0.019). The coefficient for anger intensity is 430.870, indicating that higher anger intensity is associated with an increase of approximately 431 new deaths ($p = 0.000$). Joy intensity and pred_stance do not show significant effects on new deaths.

4.2 Analysis

The research findings highlighted significant differences in public sentiment and emotional responses towards mask-wearing between Thailand and the USA. In Thailand, there was strong and consistent support for mask-wearing across all major news publishers, with fear being the predominant emotion, reflecting a cultural emphasis on collective well-being and compliance with public health directives. Conversely, the USA showed a more polarized stance, with CNN viewers being more supportive and Fox News viewers more opposed to mask-wearing. Anger was the most pronounced emotion in the USA, highlighting the contentious nature of the debate. These differences underscore the need for culturally tailored public health communication strategies.

Chapter 5
Conclusion

Conclude your Work

The findings of this project reveal significant differences in public sentiment and emotional responses towards mask-wearing between Thailand and the USA. In Thailand, there is strong and consistent support for mask-wearing across all major news publishers, with fear being the predominant emotion. This reflects a cultural emphasis on collective well-being and compliance with public health directives. Conversely, the USA displayed a more polarized stance towards mask-wearing, with CNN viewers being more supportive and Fox News viewers more opposed. Anger was the most pronounced emotion in the USA, highlighting the contentious nature of the mask-wearing debate.

Throughout this research project, I have learned to utilize advanced NLP techniques and fine-tune state-of-the-art models for sentiment and emotion analysis. Specifically, I gained practical experience in data collection using the YouTube API and preprocessing data, including translating Thai comments to English for analysis. I became proficient in using the GEMMA 2 model and EmoLLama-chat-7b for stance and emotion detection, respectively. I also learned about parameter-efficient fine-tuning using Low-Rank Adaptation (LoRA) and managing model training on large datasets with limited computational resources. Furthermore, I developed skills in evaluating model performance using metrics such as the macro F1 score and interpreting confusion matrices to identify areas for improvement. Analyzing public sentiment and emotional responses provided insights into the cultural and media influences on public opinion during health crises, which are crucial for effective public health communication.

Expectation

During my research internship, I gained invaluable insights into the life of a researcher and the experience of being a master’s student. Immersing myself in a research environment allowed me to understand the dynamics and expectations of conducting independent research.

I learned the basic steps of conducting research, including formulating research questions, collecting and analyzing data, and presenting findings. This internship provided an opportunity to learn how to conduct self-guided research, a significant departure from the structured learning environment of regular classes. I gained a deeper understanding of Natural Language Processing (NLP) and expanded my knowledge in data science.

- Additionally, this experience allowed me to explore my passion for the data science field, helping me determine if it aligns with my long-term career goals. Applying theoretical knowledge to practical research fostered skills that are crucial for a successful career in data science and research.

Benefits

Benefits to yourself

- Gained hands-on experience in conducting independent research.
- Improved technical skills in Natural Language Processing (NLP) and data science.
- Acquire practical communication skills.

Benefits to company

- Provided valuable insights into public sentiment and emotional responses to mask-wearing.
- Collected and analyzed a unique dataset of YouTube comments on mask-wearing, providing a new resource for future studies.
- Helped explore new ideas like fine-tuning large language models (LLMs).
- Demonstrated the effectiveness of using YouTube data instead of traditional sources like tweets.

Benefits to university

- Strengthened the partnership between my university and NAIST.

Swot analysis

i. **Strengths**

- Eager to Learn
- Work Hard
- Responsible to Complete Task on time
- Always Show up

ii. **Weakness**

- Hard to communicate technical concepts in English.
- Sometimes lacks confidence in presenting project findings to a broader audience, affecting the delivery of presentations.

iii. **Opportunities**

- Opportunity to get a brief taste of what advanced studies (Master's/Ph.D.) are like, to deepen research skills.
- Opportunity to improve knowledge in data science and NLP.

iv. **Threats**

- Communication barriers due to limited proficiency in Japanese.
- In a researcher's life, unlike in a bachelor's program, there are no fixed classes or structured guidance; there is no one to push you or supervise closely, requiring a high level of self-motivation and responsibility to succeed.

References

[1] Davydova, V., & Tutubalina, E. (2022). SMM4H 2022 Task 2: Dataset for stance and premise detection in tweets about health mandates related to COVID-19. In Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task (pp. 216–220). Association for Computational Linguistics.

[2] Davydova, V., Yang, H., & Tutubalina, E. (2023). Data and models for stance and premise detection in COVID-19 tweets: Insights from the Social Media Mining for Health (SMM4H) 2022 shared task. Journal of Biomedical Informatics, 149, 104555.

[3] Google Cloud: Google Cloud, 2024. Vertex AI. Available: https://cloud.google.com/vertex-ai?hl=en [Accessed on: 26 June 2024].

[4] Google Developers: Google Developers, 2024. YouTube API v3 Documentation. Available: https://developers.google.com/youtube/v3/docs/?apix=true [Accessed on: 3 June 2024].

[5] Hugging Face: Hugging Face, 2024. Hugging Face. Available: https://huggingface.co/ [Accessed on: 1 June 2024].

[6] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-Rank Adaptation of Large Language Models. In International Conference on Learning Representations.

[7] Liu, Z., Yang, K., Zhang, T., Xie, Q., & Ananiadou, S. (2024). EmoLLMs: A Series of Emotional Large Language Models and Annotation Tools for Comprehensive Affective Analysis.

[8] PyTorch: PyTorch, 2024. PyTorch. Available: https://pytorch.org/ [Accessed on: 24 May 2024].

[9] Ramos, P., Ferawati, K., Liew, K., Aramaki, E., & Wakamiya, S. (2022). Emotion Analysis of Writers and Readers of Japanese Tweets on Vaccinations. In Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (pp. 95–103). Association for Computational Linguistics.

[10] Unsloth: Unsloth, 2024. Unsloth: Your AI Productivity Assistant. Available: https://unsloth.ai/ [Accessed on: 1 July 2024].

[11] WHO COVID-19 Data Dashboard: WHO, 2024. WHO COVID-19 Data Dashboard. Available: https://data.who.int/dashboards/covid19/data [Accessed on: 26 July 2024].