Aspect and Sentiment Detection in Vietnamese Text using Transfer Learning and LSTM

Abstract. Aspect detection and sentiment classification is a challenging task in the field of natural language processing (NLP), especially in Vietnamese due to its complex grammatical structure and diversity of expressions. In this study, we propose a method of using transfer learning (BERT) and LSTM to simultaneously detect aspects and sentiments on customer feedback on mobile phone products. BERT is used to extract semantic features from text, while LSTM processes these features to aspects detection and sentiment classification. These results show that our method outperforms traditional deep learning methods. The experiment results show that our model achieved an accuracy of 93.5%, F1-score of 88.7% on the overall performance of the model. For the aspects detections, our model get accuracy at 81.5% and F1-score at 89.9%. And for the sentiment classification, an accuracy of 90.6% and F1-score at 90.6% on all aspects of model. This paper also presents detailed evaluations of the model's performance, discusses its limitations, and suggests future research directions to further improve the model's results when applied in practice.

Keywords: Transfer learning, BERT, LSTM, Natural Language Processing, Aspect Based Sentiment Analysis.

1 Introduction

The growth of e-commerce has changed the way consumers make purchasing decisions, in which they rely more and more on the feedback and reviews of previous users to judge the quality of products. These responses not only help consumers make decisions, but they also provide merchants with valuable information to tailor products and services, improve customer experience, and build effective business strategies.

Stemming from the above problems, the problem of detecting aspects and emotions was born as a field in Artificial Intelligence (AI) and natural language processing (NLP) that aims to automatically determine the sentiment of a text, usually according to categories such as positive, negative or neutral in a certain aspect [1]. However, applying this aspect and sentiment detection problem to business analysis to understand user feedback is still difficult, especially for Vietnamese text, which is a language with a complex grammatical structure and sometimes implicit usage.

Although many methods have been proposed to solve this problem, previous studies have solved this problem by separating it into two small tasks: extracting opinion targets and detecting emotional polarization [2]. Traditional and modern machine learning methods have also been used to extract objective views and detect emotions simultaneously [2]. Each method has certain advantages. However, in general, these methods are still not highly accurate and need further improvement.

Recently, the development of large language models (LLMs) has opened up new opportunities in improving the accuracy and efficiency of the problem of detecting aspects and emotions in Vietnamese texts. LLM models such as BERT, GPT and other variants have demonstrated a high level of text understanding and semantic processing [3] to improve the quality of text analysis.

Therefore, in this paper, we propose a method that combines transfer learning with BERT and LSTM models to simultaneously detect aspects and classify sentiment on the Vietnamese dataset of users' responses to the phone, in order to take advantage of the power of large language models and improve the prediction accuracy on the Vietnamese documents.

Table 1. Three examples from the customer feedback dataset on mobile phone products.

Example Comment Sentences	Aspect Sentiment Label
"Just bought this device at Thegioididong Thot Not feel it's okay. The battery is strong {BATTERY#Positive}, takes good photos {CAMERA#Positive}, loudspeakers, strong Wi-Fi connection {FEATURES#Positive}, stable signal, reasonably priced {PRICE#Positive}, and the staff are very enthusiastic in advising {SER&ACC#Positive}." {GENERAL#Positive} *"Mới mua máy này Tại thegioididong thốt nốt cảm thấy ok bin trâu {BATTERY#Positive}, chụp ảnh đẹp {CAMERA#Positive}, loa nghe to bắt wf khỏe sóng ổn định {FEATURES#Positive}, giá thành vừa với túi tiền {PRICE#Positive}, nhân viên tư vấn nhiệt tình {SER&ACC#Positive}." {GENERAL#Positive}*	{CAMERA#Positive}; {FEATURES#Positive}; {BATTERY#Positive}; {PRICE#Positive}; {GENERAL#Positive}; {SER&ACC#Positive}
"The battery is poor {BATTERY#Negative}, but everything else is great {GENERAL#Positive}. Bought it on 8/3/2019 and the battery health is down to 88%. Is anyone else experiencing this? {OTHERS}" *"Pin kém {BATTERY#Negative} còn lại miễn chê {GENERAL#Positive} mua 8/3/2019 tình trạng pin còn 88% có ai giống tôi không {OTHERS}"*	{BATTERY#Negative}; {GENERAL#Positive}; {OTHERS};
"Everyone, update your software, it will help reduce battery drain {BATTERY#Neutral}. I've tried it, and everything is okay, but the fingerprint sensor isn't very responsive {FEATURES#Negative}". {GENERAL#Neutral} *"Mọi người cập nhật phần mềm lại , nó sẽ bớt tốn pin {BATTERY#Neutral}, mình đã thử rồi, mọi thứ cũng ok, nhưng vân tay ko nhạy {FEATURES#Negative}". {GENERAL#Neutral}*	{FEATURES#Negative}; {BATTERY#Neutral}; {GENERAL#Neutral};

With the advantage of reducing training costs and processing time when performing detection and classification for input. This method is aimed at practical application in e-commerce systems, customer experience management, and big data analytics platforms. This method has the potential for wide application not only in e-commerce, but also in other fields such as customer experience management and big data analyt-

ics. Our research results provide an effective solution for Vietnamese text processing and improve the quality of user feedback analysis.

Key contributions to this article include:

- Propose a combined method of BERT and LSTM for aspect detection and sentiment classification.
- Using transfer learning from large language models to improve the accuracy and efficiency of Vietnamese text analysis.
- Comprehensively evaluate the effectiveness of the model on multiple aspects and discuss limitations and propose future improvement directions.

2 Related work

Aspect-based sentiment analysis has become an important area of study in recent years, especially with the strong development of deep learning techniques and large language models. A variety of methods have been proposed to address related challenges, from traditional machine learning models to modern deep learning methods, which improve performance in aspect detection and sentiment classification.

Luc et al. (2021) studied the use of BiLSTM (Bidirectional Long Short-Term Memory) network for aspect-based sentiment analysis, but the results were not positive when the accuracy was only average, especially for Vietnamese texts [4]. This study shows that traditional deep learning models have difficulty in handling the complex and multi-meaning contexts that are typical of the Vietnamese language.

Thin et al. (2018) extended the above method by using neural networks to detect aspects for Vietnamese, thereby identifying pairs of entities and attributes expressed in the text [6]. Although the results are improved compared to traditional methods, the accuracy is still significantly lower. One of the main problems is the limitation of these methods in processing large texts and the semantic flexibility of the Vietnamese.

The advent of large language models (LLMs) such as BERT and GPT marked a major step forward in the field of natural language processing, especially in context-based and semantic text analysis. These LLM models represent a major step forward in Artificial Intelligence (AI) and especially towards the goal of human-like synthetic artificial intelligence [5]. BERT (Bidirectional Encoder Representations from Transformers), developed by Devlin et al. (2019) has demonstrated a strong ability to process the two-dimensional semantics of text and understand context more deeply than previous models [8]. Studies such as those of Mickel et al. (2019) have applied BERT to face-based sentiment analysis and have shown results that are superior to traditional deep learning methods [1]. Thin et al. (2022) proposed an effective generic multi-tasking architecture based on neural network models to solve two tasks in ABSA, which is designed to predict the entire category of aspects in question and emotional polarizations [7]. The results of this study show that the model achieves high performance in analyzing sentiment based on the aspect but has not actually achieved the expected results.

Although BERT has achieved great success in many languages, the application of BERT to Vietnamese is still limited due to the lack of large-scale training data and

specialized models for the language. To overcome this, many recent studies have combined BERT with other deep learning models, such as LSTM, to harness the power of both models in aspect detection and sentiment classification on Vietnamese texts [7]. This combined method improves accuracy by taking advantage of BERT's text-specific extraction capabilities and LSTM's data string processing capabilities.

Our research continues this development direction, focusing on the use of the BERT large language model in combination with LSTM to simultaneously solve both the task of aspect detection and sentimental classification on Vietnamese texts. Not only do we improve performance compared to traditional methods, but we also emphasize the importance of using transfer learning to enhance the generalization capabilities of the model

3 Method Propose

In this study, we used a dataset of customer feedback on mobile phone products in Vietnamese, we built a comprehensive process to achieve the goal in this study. We first select BERT. The dataset and BERT version we use are both open and free for research purposes.

Next, we use BERT to convert sentences into a format that BERT can handle. We decided to keep the sentence format unchanged without any pure text processing, in order not to lose the style of the semantic features, thereby helping the model to learn the most complex features in Vietnamese.

In each input sentence, we stipulate a maximum length of 128 words. Sentences of greater length will be truncated, while shorter sentences will be processed using the Padding technique to ensure all sentences are 128 tokens in length. Adding Padding tokens has no semantic meaning and will not affect the learning process of the model [8]. Adjusting the maximum length makes the model work better with long sentences but also increases the computational cost.

We then encode the sentences into tokens to train BERT. We use a multilingual version of bert-base-multilingual-cased. After BERT extracts the features from the text and converts them into 768-dimensional characteristic vectors, we build the LSTM architecture to process the output of the BERT. The LSTM architecture is set up with 256 hidden units and processed in 2 directions of the chain before giving the result. We add a linear layer to perform the classification of the model's outputs with the number of outputs calculated using the formula: $Number\ of\ output = n_{classes} * n_{label\_per\_class}$. Where $n_{classes}$ is the number of aspects and $n_{label\_per\_class}$ is the number of emoticons for each aspect. This means that the model makes predictions for multiple layers and multiple labels on each layer. Note that in this study we specified the output label as "None" for classes that are not in the text.

3.1 Dataset

The UIT-ViSFD dataset used in this study has been previously described in detail by Luc et al. (2021) [4]. Below, we provide a summary of the dataset's key characteris-

tics relevant to this study. This dataset was collected from a Vietnamese e-commerce website. To ensure diversity, objectivity, and accuracy in the data set, responses from 10 phone brands were collected. The annotation principle in a dataset is followed by a strict annotation process to ensure data quality. The data was divided into three separate volumes: the training (Train), development (Dev) and testing (Test) with a ratio of 7:1:2 respectively. Table 2 presents the overview statistics of the UIT-ViSFD dataset. Number of Comments is the number of comments collected, Number of Tokens is the number of words, Number of Aspects is the number of aspects extracted from comments, Average number of aspects per sentence is the number of aspects per sentence and Average length per sentence is the average length of sentences.

Table 2. Overview statistics of Train/Dev/Test sets of UIT-ViSFD dataset [4].

	Train	**Dev**	**Test**
Number of Comments	7,786	1,112	2,224
Number of Tokens	283,460	39,023	80,787
Number of Aspects	23,597	3,371	6,742
Average number of aspects per sentence	3.3	3.2	3.3
Average length per sentence	36.4	35.1	36.3

Table 3 depicts the distribution of aspects and sentiments in the Train, Dev, and Test suites. The aspects have a marked imbalance, the most noticeable being that of the GENERAL class is 6,936 data points while STORAGE has only 132. There is also a big difference between the three emotional poles. The number of sensory counts also varies greatly, with Pos accounting for 56.13%, followed by Neg at 31.70% and finally Neu at 12.17%. This dataset is unbalanced [4].

Table 3. The distribution of aspects and their sentiments of UIT-ViSFD dataset [4].

Aspect	**Train**			**Dev**			**Test**			**Total**
	Pos	**Neg**	**Neu**	**Pos**	**Neg**	**Neu**	**Pos**	**Neg**	**Neu**	
BATTERY	2,027	349	1,228	303	51	150	554	92	368	5,122
CAMERA	1,231	288	627	172	36	88	346	71	171	3,030
DESIGN	999	77	302	135	12	40	274	28	96	1,963
FEATURES	785	198	1,659	115	33	233	200	52	459	3,734
GENERAL	3,627	290	949	528	34	127	1,004	83	294	6,936
PERFORMANCE	2,253	391	1,496	327	45	210	602	116	454	5,894
PRICE	609	391	316	72	144	36	162	328	79	2,882
SCREEN	514	56	379	62	12	47	136	17	116	1,339
SER&ACC	1,401	107	487	199	13	78	199	27	167	2,678
STORAGE	59	107	21	11	1	2	18	3	6	132
Total	13,505	2,903	7,464	1,924	381	1,011	3,495	817	2,210	

This dataset presents several challenges for aspect detection and sentiment classification due to the significant class imbalance. The imbalance of data can lead to two problems, the first is that the model will learn better in large numbers of classes, but fewer classes will tend to be ignored. The second is that the model may have more difficulty in learning the characteristics of the STORAGE class, leading to the model being mispredicted for classes with too few labels, thereby making the recognition less accurate. For instance, the STORAGE aspect has only 132 instances, compared to 6,936 instances for the GENERAL aspect. Additionally, many comments contain multiple aspects in the same sentence, making it harder for the model to accurately detect and classify sentiments for each aspect.

3.2 State-of-the-art models

In this study, our overview model is built on a combination of two advanced models, BERT and LSTM, for both aspect detection and sentiment classification tasks simultaneously for a single output.

BERT architecture consists of a multi-layered two-dimensional Transformer encoder based on the original implementation described and released in the *tensor2tensor* library. The baseline BERT model we used in this study had the following parameters: *L=12, H=768, A=12*, and the total number of parameters was 110M. Where *L* is the number of layers, the hidden layer size is *H* and *A* is the number of self-attention heads [8]. BERT is a powerful linguistic model that uses an attention mechanism to learn the semantic features of text in two dimensions. The BERT model is trained on large amounts of data and can capture deep semantic relationships in text thanks to the use of a two-dimensional self-attention mechanism [8]. The BERT model works according to the "*Masked Language Model (MLM)*" mechanism during the training phase, in which some words in the sentence are obscured, and the model must predict these obscured words based on the surrounding context. In addition, BERT is also trained with the "*Next Sentence Prediction (NSP)*" mechanism to understand the relationship between sentences in the text [8]. These characteristics allow BERT to provide very powerful semantic characteristic vectors, which can be used in a variety of natural language processing tasks, including sentiment analysis, information extraction, and many others.

LSTM architecture was first introduced by S Hochreiter el al. [9] is designed to process and predict sequential data, which is especially useful when data has a long-term dependency such as text. LSTM has a special structure with the following main steps:

This Gate: decides how much information from the previous state C_{t-1} will be retained. The Forget Gate receives input from the hidden state of the previous step h_{t-1} and the current input x_t then applies a sigmoid function to calculate the value f_t:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

The value f_t is in the range [0,1], where 0 means completely forget and 1 means fully retain.

Input Gate: This gate decides what new information will be added to the Cell State C_t. There are two main parts: First, a sigmoid function decides which values need to be updated, resulting in i_t:

$$\mathrm{i_t} = \sigma(\mathrm{W_i} \cdot [\mathrm{h_{t\text{-}1}}, \mathrm{x_t}] + \mathrm{b_i})$$

Then, a *tanh* layer creates a new vector of values $\widetilde{C_t}$:

$$\widetilde{C_t} = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Update Cell State: The Cell State is updated by combining the information that needs to be forgotten and the new information:

$$\mathrm{C_t} = \mathrm{f_t} \cdot \mathrm{C_{t-1}} + \mathrm{i_t} \cdot \widetilde{\mathrm{C_t}}$$

Here, the multiplication · is element-wise multiplication.

Output Gate: Finally, to determine the output for the current hidden state h_t, the LSTM uses the output gate. First, a sigmoid function calculates which part of the Cell State will be output as $\mathrm{o_t}$:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

Then, the current hidden state h_t is calculated by multiplying o_t by the tanh of the new Cell State C_t:

$$\mathrm{h_t} = \mathrm{o_t} \cdot \tanh(\mathrm{C_t})$$

3.3 Our model performance evaluation method

In this study, we evaluate the model from various dimensions to examine its performance from a holistic to a detailed perspective. First, we assess the overall performance of the model to simultaneously check its ability to detect and classify aspects and sentiments. Next, we evaluate the model's performance in detecting aspects for each individual aspect as well as across all aspects. Similarly, we assess the model's sentiment classification performance for each aspect and across all aspects. A detailed evaluation from various dimensions is necessary to gain a deeper understanding of the model's capabilities, as well as to identify its strengths and limitations in specific aspects. The performance metrics used in the evaluation include Accuracy, Precision, Recall, and F1-score, which help measure the model's ability to detect and classify sentiments.

4 Experimental results

In this study, we neglected to detect the aspect and categorize the emotions for the OTHERS class because they do not express their own emotions in the paragraph. Figure 1 shows the model's loss index during training starting at 0.35 and converging at 0.01 after going through 40 epochs. The descending loss graph shows that the model has learned well-learned features that can distinguish more accurately over multiple epochs.

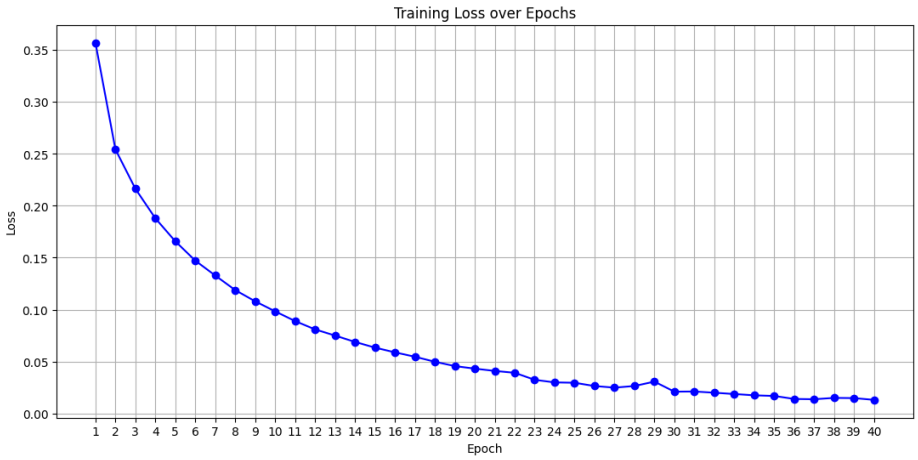

Fig. 1. The loss metric chart of the model over 40 epochs of training.

Table 4 presents the results for the overall performance of the model, the level of aspect recognition, and the ability to classify sentiment based on the aspect. The model performance evaluation metrics include Accuracy, Precision, Recall, and F1-score.

The overall performance of the model achieved an Accuracy of 93.5% and an F1-score of 88.7%. This result demonstrates that the model is capable of accurately classifying aspects and sentiments, with a balance between Precision and Recall of 88.0% and 88.5%, respectively.

For the model's performance in aspect detection, it achieved high results with an Accuracy of 81.5% and an F1-score of 89.8%. This indicates that the model is capable of detecting aspects with high accuracy, with Precision and Recall scores of 88.1% and 91.6%, respectively.

In sentiment classification, the model achieved an Accuracy and F1-score of 90.6% each. The high Precision score of 90.8% and Recall of 90.6% indicate that the model is effective at classifying sentiment, with minimal misses or misclassifications.

Table 4. Evaluation of Overall Performance, Aspect Detection, and Sentiment Classification Metrics for the Model.

	Overall Performance	**Aspect Detection**	**Sentiment Detection**
Accuracy	93.5%	81.5%	90.6%
Precision	88.0%	88.1%	90.8%
Recall	88.5%	91.6%	90.6%
F1-score	88.7%	89.9%	90.6%

Table 5 shows the results of evaluating the performance of the model in terms of recognition on each aspect and the classification of emotions for each aspect.

Table 5. Kết quả các chỉ số đánh giá hiệu suất (%) nhận diện khía cạnh và hiệu suất phân loại cảm tính của mô hình trên từng khía cạnh.

Aspect	Aspect Detection				Sentiment Detection			
	Acc	**Pre**	**Recall**	**F1**	**Acc**	**Pre**	**Recall**	**F1**
BATTERY	96.7	94.8	98.1	96.4	90.0	91.0	90.0	91.0
CAMERA	97.4	93.0	97.6	95.2	94.0	95.0	94.0	94.0
DESIGN	95.5	87.5	87.9	97.7	94.0	94.0	94.0	94.0
FEATURES	91.2	82.9	91.4	86.9	88.0	89.0	88.0	89.0
GENERAL	85.4	86.3	90.9	88.5	83.0	83.0	83.0	83.0
PERFORMANCE	90.1	90.0	91.4	90.7	84.0	84.0	84.0	84.0
PRICE	95.0	89.4	91.2	90.3	90.0	90.0	90.0	90.0
SCREEN	96.1	81.7	86.6	84.1	95.0	95.2	95.0	94.9
SER&ACC	90.4	81.4	82.9	82.2	89.0	89.0	89.0	89.0
STORAGE	99.3	75.0	66.6	70.5	99.0	99.0	99.0	99.0

The results in Table 5 indicate that the model performs well in aspect detection and sentiment classification. For the model's aspect recognition ability, based on Accuracy and F1-score metrics, we can see that aspects such as BATTERY, CAMERA, FEATURES, PERFORMANCE, and PRICE are well-recognized by the model, with Accuracy and F1-score metrics all exceeding 90%. Aspects with stable Accuracy or F1-score metrics around 80% include DESIGN, GENERAL, SCREEN, and SER&ACC. For the STORAGE aspect, although the Accuracy is very high at 99.3%, the F1-score is only 70.5% and Recall is 66.6%. This suggests that the model sometimes overlooks the STORAGE aspect and mainly focuses on aspects with a larger number of samples.

In terms of sentiment classification, the model performs well on aspects such as BATTERY, CAMERA, DESIGN, PRICE, SCREEN, and STORAGE, with Accuracy exceeding 90%. On the other hand, aspects like FEATURES, GENERAL, PERFORMANCE, and SER&ACC have both Accuracy and F1-score metrics above 80%, with FEATURES and SER&ACC results approaching 90%. Overall, the model has achieved excellent performance in sentiment classification, showing impressive results across all aspects.

5 Discusion

Overall, the model in this study has achieved good performance in aspect detection and sentiment classification on text, outperforming traditional deep learning methods such as BiLSTM or neural networks. Specifically, our method achieved F1-scores of 89.9% for aspect detection and 90.6% for sentiment classification, higher than the BiLSTM-based method used by Luc et al., which achieved F1-scores of 84.48% for aspect detection and 63.06% for sentiment classification. Our results also surpass the

neural network method employed by Thin et al., which had F1-scores of 78.66% for aspect detection and 73.16% for sentiment classification. Compared to newer methods, our model performs similarly to the BERT-based approach used by Mickel et al., and shows superior results compared to earlier traditional methods.

However, there are two issues that need to be addressed to improve the model. First, the Accuracy score in aspect detection indicates that the model needs improvement to better identify all aspects in the text. Second, although the Accuracy for detecting the STORAGE aspect is 99.3%, the Precision, Recall, and F1-score are only around 70.0%. This may be due to class imbalance in the data, as the STORAGE class has the fewest samples, making it more challenging for the model to learn to identify this class and leading to a greater focus on classes with more samples. Our results are consistent with the method used by Luc et al., which also showed that the model struggled with detecting the STORAGE aspect.

6 Conclusion

This study proposed a method using transfer learning by combining the large language model BERT with LSTM to simultaneously perform two tasks: aspect detection and sentiment classification in a single task. This differs from traditional methods, which separate the task into two distinct steps: detecting aspects first, followed by sentiment classification.

The results of the study show that our proposed method outperforms traditional methods when applied to the same dataset. Furthermore, this method demonstrates higher stability when working with Vietnamese, which has many complex grammatical features.

However, this study has some limitations. First, our model struggles with handling sentences where there is an imbalance in the number of samples between classes. To address this, we propose using data augmentation techniques or collecting additional data from similar sources. Second, the dataset used in this study is limited in scale and diversity. While the results on this dataset are promising, the model's generalization ability on other data or from different sources may be affected. Therefore, a future research direction is to expand the dataset and test the model on various datasets and languages.

Finally, another promising avenue to explore is the use of more advanced modern language models combined with traditional methods to leverage the strengths of both approaches. We also recommend further research into optimizing model parameters to improve performance even more.

References

1. Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. 2019. Aspect-Based Sentiment Analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 187–196, Turku, Finland. Linköping University Electronic Press.

2. Mai, L., Le, B. (2018). Aspect-Based Sentiment Analysis of Vietnamese Texts with Deep Learning. In: Nguyen, N., Hoang, D., Hong, TP., Pham, H., Trawiński, B. (eds) Intelligent Information and Database Systems. ACIIDS 2018. Lecture Notes in Computer Science (), vol 10751. Springer, Cham. https://doi.org/10.1007/978-3-319-75417-8_14.
3. Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. *arXiv preprint arXiv:2402.06196*.
4. Luc Phan, L., Huynh Pham, P., Thi-Thanh Nguyen, K., Khai Huynh, S., Thi Nguyen, T., Thanh Nguyen, L., ... & Van Nguyen, K. (2021). Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence. In *Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part II 14* (pp. 647-658). Springer International Publishing.
5. y Arcas, B. A. (2022). Do large language models understand us?. *Daedalus*, *151*(2), 183-197.
6. Van Thin, D., Nguye, V. D., Van Nguyen, K., & Nguyen, N. L. T. (2018, November). Deep learning for aspect detection on vietnamese reviews. In *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)* (pp. 104-109). IEEE.
7. Van Thin, D., Le, L. S., Nguyen, H. M., & Nguyen, N. L. T. (2022). A joint multi-task architecture for document-level aspect-based sentiment analysis in vietnamese. *IJMLC*, *12*(4).
8. Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
9. Hochreiter, S. (1997). Long Short-term Memory. *Neural Computation MIT-Press*.