Gen-mABSA-T5: A Multilingual Zero-Shot Generative Framework for Aspect-Based Sentiment Analysis

Shabrina Akter Shahana
Dept. of CSE, Jahangirnagar University
`shabrina.stu2018@juniv.edu`

Nuzhat Nairy Afrin
Dept. of CSE, Jahangirnagar University
`nuzhat.stu2019@juniv.edu`

Md. Musfique Anwar
Dept. of CSE, Jahangirnagar University
`manwar@juniv.edu`

Israt Jahan
Dept. of CSE, Jahangirnagar University
`israt@juniv.edu`

Abstract

Aspect-Based Sentiment Analysis (ABSA) identifies sentiments toward specific aspects of an entity. While progress has been substantial for high-resource languages such as English, low-resource languages like Bangla remain underexplored due to the limited availability of annotated data and linguistic challenges. We propose Gen-mABSA-T5, a multilingual zero-shot generative framework for ABSA based on Flan-T5, incorporating prompt engineering and Natural Language Inference (NLI). Without task-specific training, Gen-mABSA-T5 achieves state-of-the-art zero-shot accuracy of 61.56% on the Large Bangla Corpus, 73.50% on SemEval Laptop, and 73.56% on SemEval Restaurant outperforms both English and Bangla task-specific models in the zero-shot settings. It delivers reasonable performance against very large general-purpose models on both English and Bangla benchmarks. These results highlight the effectiveness of generative, zero-shot approaches for ABSA in low-resource and multilingual settings.

1 Introduction

Bangla is the seventh most spoken language worldwide (Shammi et al., 2023), yet Aspect-Based Sentiment Analysis (ABSA) for Bangla remains underexplored due to scarce annotated data and language-specific complexities. ABSA, a fine-grained extension of sentiment analysis, identifies sentiments toward specific aspects of an entity rather than providing only an overall polarity. For example, in "The phone's camera is excellent, but the battery is weak," the sentiment is positive for the *camera* but negative for the *battery*. This level of detail makes ABSA valuable in applications such as e-commerce, service reviews, and social media monitoring, where organizations require targeted insights into user opinions.

Despite its importance, ABSA faces significant challenges. Traditional supervised approaches rely on large domain-specific datasets, which are rarely available for low-resource languages like Bangla. Even multilingual transformer models such as BERT and RoBERTa, though effective in English, underperform in Bangla due to limited exposure to its morphology and syntax. Furthermore, ABSA is complicated by sentiment ambiguity and domain variation, highlighting the need for approaches that minimize reliance on annotated data while generalizing across domains and languages.

Zero-shot ABSA addresses these issues by enabling models to identify aspects and classify their sentiment without task-specific training. Instead of requiring labeled corpora for each new domain, zero-shot models leverage pre-trained knowledge and adapt through mechanisms such as prompt engineering and natural language inference (NLI). This capability is particularly valuable for Bangla, where annotated resources are scarce, and for multilingual contexts, where scalability is essential.

This work introduces Gen-mABSA-T5, a multilingual generative zero-shot framework for ABSA built on Flan-T5. By integrating prompt engineering and with the framework's NLI capabilities, Gen-mABSA-T5 performs aspect sentiment classification in Bangla and English without domain-specific supervision. Evaluations in SemEval 2014 (English) (Pontiki et al., 2014) and the Large Bangla Corpus demonstrate that Gen-mABSA-T5 achieves state-of-the-art zero-shot accuracy, outperforming both English and Bangla task-specific models in zero-shot settings. It delivers reasonable performance against very large general-purpose models on both English and Bangla benchmarks, underscoring the model's adaptability to both high and low-resource settings.

The contributions of this work are as follows:

Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025), pages 142–150

- We propose Gen-mABSA-T5, a multilingual zero-shot generative framework for ABSA that does not require task-specific training data.
- We conducted an extensive evaluation on English and Bangla datasets, achieving state-of-the-art zero-shot results for Bangla and reasonable zero-shot performance for English.
- We demonstrate the effectiveness of prompt engineering and NLI for cross-lingual ABSA, highlighting the scalability of zero-shot generative approaches for low-resource languages.

2 Literature Review

The field of Aspect-Based Sentiment Analysis (ABSA) has witnessed significant advances in recent years, driven by developments in natural language processing and deep learning. Existing research covers a broad spectrum, from rule-based and feature-based approaches to state-of-the-art transformer models. However, most prior work relies heavily on large annotated datasets, which limits applicability in low-resource languages. Recent trends focus on multilingual and zero-shot approaches to overcome these limitations. We reviewed relevant studies, identifying strengths, limitations, and gaps that motivate the proposed Gen-mABSA-T5 framework.

SA has evolved from lexicon and rule-based systems to statistical models, deep learning, and transfer learning. In English, early approaches used sentiment lexicons (e.g., SentiWordNet) and parsers, with SemEval(2014) establishing benchmarks. The advent of BERT and related transformers marked a paradigm shift, while zero-shot and few-shot learning extended applicability in low-resource settings (Li and Xiang., 2020; Pathan and Prakash., 2022).

Bangla SA has developed more slowly due to limited resources. Early studies relied on translated or code-mixed datasets (Dey and Sarker., 2019; Mahtab et al., 2018), lexicon-based methods (Rabeya and Sattar., 2022), and multilingual models (Alam and Kamal., 2024). More recent work has applied deep learning and multilingual transformers (Agüero-Torales and López-Herrera., 2021), although high-quality annotated corpora remain scarce.

ABSA refines SA by assigning sentiment at the aspect level. Early English ABSA employed lexicon-based methods and dependency parsing, later replaced by pre-trained models, unified pipelines, and generative frameworks such as T5 and BART (Zhang and Lam., 2021; Huan and Guo., 2022). In Bangla, research began with traditional ML baselines (Rahman and Dey., Data 3, no. 2 (2018; Rahman and Kumar Dey, 2018), progressed to deep learning with CNN, LSTM, and RNN architectures (Mahfuz Ahmed Masum and Islam., 2020; Md Morshedul Islam and Myneddin., 2023), and has recently adopted LLMs (Shihab Ahmed and Mridha, 2024; Md Akash Rahman and Andersson., 2024; Junayed Hossain and Monir., 2024), improving performance on complex tasks such as implicit aspect detection.

Zero-shot ABSA has emerged as a promising alternative to supervised approaches. In English, defining ABSA as a Natural Language Inference (NLI) task enabled models such as BERT and T5 to perform aspect extraction and sentiment classification without domain-specific training (Shu et al., 2022). More recently, GPT-3 and GPT-4 have demonstrated robust zero-shot performance across unseen domains. For Bangla, progress is limited by morphological complexity and resource scarcity, though multilingual models (mBERT, XLM-R) enable some transfer from English with mixed results.

Comparing approaches, traditional ML offers interpretability on small datasets, deep learning improves accuracy but requires large corpora, while LLMs provide advanced contextual understanding and strong performance in multilingual and zero-shot settings.

In summary, English SA and ABSA have matured through robust datasets and methodologies, while Bangla research remains constrained by limited resources. Recent advances in deep learning and LLMs, however, have opened promising directions for Bangla sentiment analysis and ABSA. These developments underscore the importance of scalable, zero-shot approaches capable of bridging the gap between high- and low-resource languages.

3 Proposed Methodology: Gen-mABSA-T5

We introduce Gen-mABSA-T5, a multilingual zero-shot generative framework for aspect-based sentiment analysis (ABSA), which addresses challenges in low-resource languages such as Bengali, where annotated data is limited. Unlike traditional ABSA models that depend on large labeled datasets, Gen-mABSA-T5 leverages zero-shot learning, prompt engineering, and a generative transformer architecture to perform ABSA without task-specific training.

3.1 System Overview

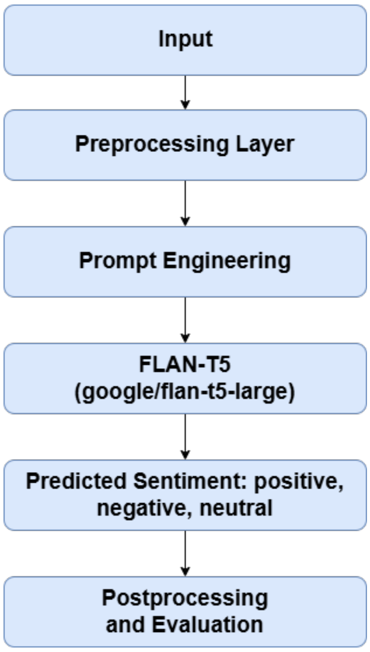

Figure 1: Gen-mABSA-T5 System Overview

The framework is based on Flan-T5 and integrates prompt engineering with NLI to classify sentiment toward a given aspect in a fully generative manner. Input consists of a sentence and an explicitly provided aspect; output is a single sentiment label (positive, negative, or neutral). No fine-tuning or separate NLI head is used. NLI is achieved through carefully designed prompts that take advantage of Flan-T5's pre-training on inference tasks.

3.2 Model Architecture

Figure 2 shows the full end-to-end pipeline of Gen-mABSA-T5. The model receives a raw sentence together with an explicitly annotated aspect term (no aspect extraction is performed). The sentence first passes through a preprocessing layer that removes all emojis, deletes punctuation marks used in both English and Bangla (including the Bangla full stop , double full stop , curly quotes , and zero-width joiners that can break compound Bangla words), and collapses multiple spaces into one. This cleaned sentence is inserted into a prompt style: For example, in English, "Sentence: [text]. Aspect: [aspect]. Classify sentiment toward the aspect: positive, negative, or neutral." and in Bangla the exact Bangla translation of the same instruction. The completed prompt is fed directly to the publicly available `Google/Flan-t5-large` model under zero-shot inference with fixed settings (temperature 0, maximum 10 new tokens). The model generates the word "positive", "negative", or "neutral", which is then converted to the integer labels used for evaluation (positive = 2, negative = 0, neutral = 1).

3.3 Zero-Shot and NLI Framing

Zero-shot ABSA addresses these issues by enabling models to identify aspects and classify their sentiment without task-specific training. Instead of requiring labeled corpora for each new domain, zero-shot models leverage pre-trained knowledge and adapt through mechanisms such as prompt engineering and NLI. This capability is particularly valuable for Bangla, where annotated resources are scarce, and for multilingual contexts, where scalability is essential.

Gen-mABSA-T5 performs aspect sentiment classification in Bangla and English without domain-specific supervision. The framework frames ABSA as an NLI task within a generative paradigm: the input sentence serves as the premise, while a hypothesis template (The sentiment toward [aspect] is [positive/negative/neutral]) is constructed via a prompt. Flan-T5 then generates a prediction, which is directly mapped to positive, negative, or neutral. No separate NLI head or classifier is used. The reasoning of NLI emerges from Flan-T5's pre-training on datasets such as MNLI and SNLI, activated only through structured instructions. This design ensures full zero-shot operation with no fine-tuning.

3.4 Prompt Engineering and Ablation Studies

By integrating prompt engineering and NLI, Gen-mABSA-T5 performs aspect sentiment classification in Bangla and English without domain-specific supervision. Prompts were iteratively designed using a held-out validation split (10% of each training set). We tested 10 English and 10 Bangla variants, including concise, NLI-style, structured output, etc. formats. The best-

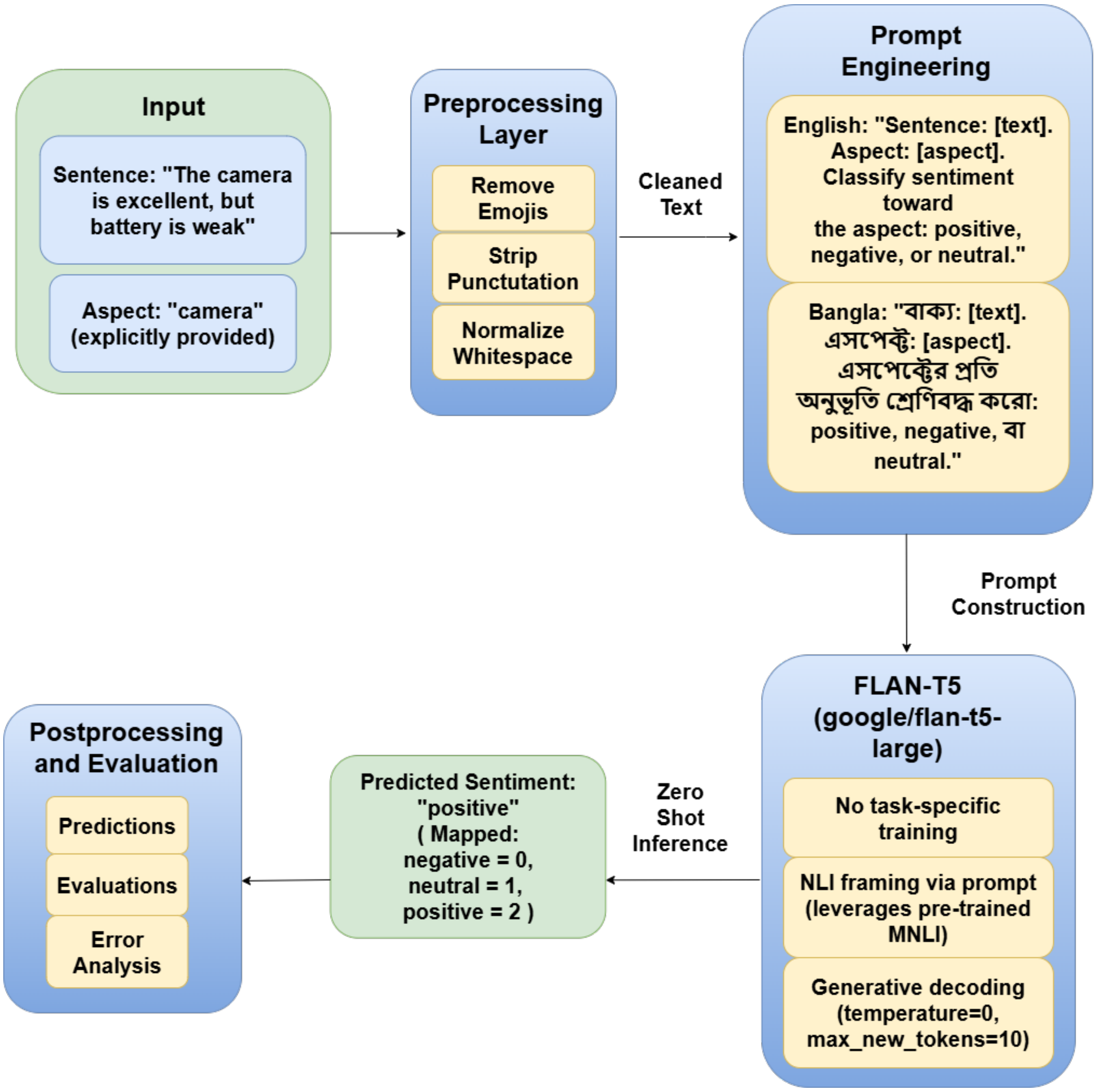

Figure 2: End-to-end pipeline of Gen-mABSA-T5

performing prompts are the following concise and explicit prompts:

> You are a sentiment analysis expert.
> Sentence: [sentence].
> What is the sentiment towards [aspect]?
> Choose only one: positive, negative, or neutral.

For Bengali, prompts are adapted to its morphology and syntax:

> তুমি একজন অনুভূতি বিশ্লেষণ বিশেষজ্ঞ।
> বাক্য: [sentence]
> [aspect]-এর প্রতি অনুভূতি কী?
> শুধুমাত্র একটি বেছে নাও: positive, negative, অথবা neutral।

Longer prompts and domain context reduced performance (-13.1% avg. accuracy), likely due to noise. Strict, concise instructions with explicit aspect focus yielded optimal zero-shot generalization. See Table 1.

3.5 Aspect Handling

The model assumes aspects are pre-provided from dataset annotations and does not perform automatic aspect extraction. Only explicit, single-word or short-phrase aspects are processed; implicit and multi-word aspects are out of scope. For each inference, one aspect is paired with the full sentence in the prompt. This design follows the standard evaluation protocol of SemEval 2014 (Pontiki et al., 2014) and the Bangla corpus, where

Prompt Style	SemEval 2014 Laptop	SemEval 2014 Restaurant	Bangla Corpus	Avg. Acc.
NLI-style Concise Explicit	0.735	0.736	0.616	0.696
Confidence Aware	0.710	0.728	0.625	0.688
Structured Output	0.716	0.721	0.612	0.683
Context Enhanced	0.619	0.594	0.508	0.574

Table 1: Ablation Study Evaluating the Impact of Different Prompt Designs on Gen-mABSA-T5's Accuracy.

aspect terms are given.

3.6 Implementation and Reproducibility

We implement Gen-mABSA-T5 using Google/Flan-t5-base (250M parameters) via Hugging Face Transformers. Inference is deterministic temperature=0.0, maximum new tokens is 10, batch size 32. The output is parsed by case-insensitive matching of positive, negative, or neutral. Text pre-processing removes emojis, normalizes punctuation, and collapses whitespace. Baselines (BERT-Multilingual, RoBERTa-Twitter, XLM-RoBERTa, Bangla-BERT) use sequence classification with 3 labels and weighted metrics. Zero-shot requires no training; few-shot samples 50 examples per aspect; fine-tuning runs 1500 steps (batch size 8, weight decay 0.01).

The SemEval 2014 Laptop and Restaurant (Pontiki et al., 2014) were split into train-test as the official distribution. The Large Bangla ABSA Corpus is split into an 80-20 ratio stratified by aspect.

The experiments were carried out on the Google NVIDIA T4 GPU (16 GB of VRAM) with 12 GB of RAM, using PyTorch and Python. All seeds were fixed at 42.

3.7 Contributions

Gen-mABSA-T5 advances ABSA by enabling multilingual support, including for low-resource languages, through zero-shot learning, reducing reliance on labeled data. Its generative approach enhances flexibility and performance in resource-limited settings, with applications in customer feedback and social media analysis. Gen-mABSA-T5 integrates NLI, multilingual transformers, and prompt engineering for scalable zero-shot ABSA.

4 Experimental Setup

We evaluated state-of-the-art sentiment analysis models for Aspect-Based Sentiment Analysis (ABSA) across diverse linguistic and domain contexts, focusing on zero-shot, few-shot, and fully fine-tuned settings. Evaluations use benchmark datasets: SemEval 2014 Task 4 (Restaurant and Laptop domains) for English, and the Large Bangla Corpus for Bangla. We assessed BERT-Multilingual, RoBERTa-Twitter, XLM-RoBERTa, Bangla-BERT, and our proposed Gen-mABSA-T5, a zero-shot multilingual ABSA model designed for low-resource languages like Bangla. The results demonstrate Gen-mABSA-T5's effectiveness, particularly in zero-shot scenarios, while identifying challenges and opportunities in multilingual ABSA.

4.1 Dataset

This section outlines the datasets used for evaluating Aspect-Based Sentiment Analysis (ABSA) models, covering English and Bangla to enable robust cross-linguistic and cross-domain analysis. For English, the SemEval 2014 Task 4 datasets, standard ABSA benchmarks, include the Restaurant domain (covering food, service, ambiance) and Laptop domain (e.g., battery life, screen quality), with sentences annotated for aspects and sentiment polarities (positive, negative, neutral). For Bangla, we utilized the Large Bangla Corpus spanning multiple domains. These datasets vary in domain and annotation quality, supporting comprehensive evaluation in low-resource settings.

4.2 Evaluation Metrics

We evaluated the metrics used for assessing Aspect-Based Sentiment Analysis (ABSA) models: Accuracy and F1-Score, selected to assess performance and address class imbalances. Accuracy measures the proportion of correct sentiment predictions (positive, negative, neutral) relative to all predictions, suitable for balanced datasets. The F1-Score, the harmonic mean of precision and recall, balances false positives and negatives, ideal for unbalanced datasets.

Model	Zero-shot		Few-shot		Fine-tuned	
	Acc	F1	Acc	F1	Acc	F1
BERT-Multilingual	0.244	0.321	0.774	0.777	0.795	0.792
RoBERTa-Twitter	0.718	0.733	0.803	0.802	0.816	0.815
XLM-RoBERTa	0.709	0.726	0.782	0.783	0.808	0.807
Bangla-BERT	0.295	0.303	0.427	0.287	0.509	0.441

Table 2: Performance on SemEval 2014 Laptop Corpus.

Model	Zero-shot		Few-shot		Fine-tuned	
	Acc	F1	Acc	F1	Acc	F1
BERT-Multilingual	0.170	0.224	0.783	0.786	0.778	0.778
RoBERTa-Twitter	0.686	0.696	0.796	0.798	0.791	0.790
XLM-RoBERTa	0.696	0.703	0.770	0.771	0.793	0.796
Bangla-BERT	0.571	0.419	0.576	0.421	0.576	0.421

Table 3: Performance on SemEval 2014 Restaurant Corpus.

4.3 Baseline Models

This subsection describes the baseline models evaluated for ABSA tasks, comprising transformer-based models, including one specifically pre-trained for Bangla. Each model is tested in zero-shot, few-shot, and fully fine-tuned settings to evaluate adaptability and performance across diverse datasets. The models include:

BERT-Multilingual: A multilingual BERT model pre-trained for sentiment analysis, suited for cross-lingual ABSA (nlptown/bert-base-multilingual-uncased-sentiment).

RoBERTa-Twitter: A RoBERTa model fine-tuned on Twitter data, effective for sentiment analysis in informal text `cardiffnlp/twitter-roberta-base-sentiment`.

XLM-RoBERTa: A cross-lingual RoBERTa model supporting multilingual sentiment analysis (cardiffnlp/twitter-xlm-roberta-base-sentiment).

Bangla-BERT: A BERT model pre-trained for Bangla, designed for low-resource language ABSA (sagorsarker/bangla-bert-base).

4.4 Evaluation Conditions

Models are evaluated in three settings to assess adaptability and performance: zero-shot, using pre-trained weights and prompts without task-specific training; few-shot, fine-tuned with approximately 10-100 labeled examples per sentiment class (positive, negative, neutral); and fully fine-tuned, trained on the entire labeled dataset for optimal supervised performance.

5 Results and Discussion

This section reports the performance of ABSA models in zero-shot, few-shot, and fully fine-tuned settings on SemEval 2014 Laptop and Restaurant datasets (English) and the Large Bangla ABSA Corpus. Metrics include accuracy and F1-score. The focus is to analyze the performance of the task-specific models and general-purpose LLMs in the ABSA task in contrast to the Gen-mABSA-T5, particularly in zero-shot settings.

5.1 Performance on SemEval 2014 Laptop Corpus

Table 2 summarizes the performance of baseline transformer models (BERT-Multilingual, RoBERTa-Twitter, XLM-RoBERTa, and Bangla-BERT) on the SemEval 2014 Laptop dataset in zero-shot, few-shot, and fine-tuned settings. In zero-shot, RoBERTa-Twitter achieves the highest accuracy (71.79 %) and F1 score (73.34 %), leveraging its pre-training on sentiment-rich Twitter data, while BERT-Multilingual lags at 24.36 % accuracy due to limited domain adaptation. Few-shot learning boosts performance across models, with RoBERTa-Twitter reaching 80.34 % accuracy, demonstrating the value of minimal labeled data for refinement. Fine-tuning yields further gains, peaking at 81.62 % accuracy for RoBERTa-Twitter, highlighting the effectiveness of full supervision on high-resource English data.

Our proposed framework achieves 73.50 % accuracy and 68.34 % F1 in zero-shot setting (Table 5), outperforming most baselines except

Model	Zero-shot		Few-shot		Fine-tuned	
	Acc	F1	Acc	F1	Acc	F1
BERT-Multilingual	0.237	0.347	0.945	0.951	0.947	0.943
RoBERTa-Twitter	0.539	0.688	0.926	0.922	0.927	0.927
XLM-RoBERTa	0.019	0.018	0.930	0.926	0.927	0.927
Bangla-BERT	0.615	0.641	0.615	0.662	0.863	0.850

Table 4: Performance on Large Bangla ABSA Corpus.

Model	Model Type	SemEval 2014 Laptop		SemEval 2014 Restaurant		Bangla Corpus	
		Acc	F1	Acc	F1	Acc	F1
BERT-Multilingual	Task-Specific	0.244	0.321	0.170	0.224	0.237	0.347
RoBERTa-Twitter	Task-Specific	0.718	0.733	0.686	0.696	0.539	0.688
XLM-RoBERTa	Task-Specific	0.709	0.726	0.696	0.703	0.002	0.002
Bangla-BERT	Task-Specific	0.295	0.303	0.571	0.419	0.615	0.614
gpt-oss-20b	General-Purpose	0.855	0.837	0.817	0.787	0.784	0.812
kimi-k2-instruct	General-Purpose	0.842	0.817	0.804	0.757	0.860	0.872
llama-3.1-8b-instant	General-Purpose	0.718	0.738	0.707	0.717	0.614	0.616
Gen-mABSA-T5	General-Purpose	0.735	0.683	0.736	0.668	0.616	0.628

Table 5: Zero-shot performance of general-purpose and task-specific models on ABSA.

RoBERTa Twitter and XLM-RoBERTa in task-specific pre-trained models. This underscores its generative zero-shot strength via prompt engineering and NLI framing without fine-tuning. Compared to general-purpose LLMs such as GPT-OSS-20B (20B parameters) (85.47 % accuracy with our designed prompt, Kimi-K2-Instruct (1T parameters) (84.19 %), and Llama-3.1-8B-Instant (8b parameters) (71.8%), our framework is commendable and lightweight.

These results demonstrate that prompt-engineered Flan-T5 can serve as a robust zero-shot baseline for English ABSA, approaching the performance of domain-specific pre-trained models without any task-specific adaptation. The gap to fine-tuned systems (approximately 5–8% F1) is expected and aligns with the zero-shot paradigm's goal of multilingual scalability rather than peak in-domain accuracy.

5.2 Performance on SemEval 2014 Restaurant Corpus

As shown in Table 3, baselines on the SemEval 2014 Restaurant dataset exhibits similar trends. XLM-RoBERTa leads in zero-shot (69.63 % accuracy, 70.28 % F1), benefiting from multilingual pre-training, while Bangla-BERT underperforms (57.07 %) due to language mismatch. Few-shot improvements are notable, with RoBERTa-Twitter at 79.58 % accuracy, and fine-tuning pushes XLM-RoBERTa to 79.32 %. These results indicate domain-specific challenges, such as varied aspect granularity in restaurant reviews.

Our framework attains 73.56 % accuracy and 66.81 % F1 in the zero-shot setting (Table 5), surpassing all task-specific models and Llama-3.1-8b-instant (70.7% accuracy) in general purpose model. Benchmarking against general-purpose LLMs, GPT-OSS-20B achieves 81.68 % accuracy and Kimi-K2-Instruct 80.37 % using our designed NLI-style concise and explicit prompt.

5.3 Performance on Large Bangla ABSA Corpus

On the Large Bangla Corpus, Gen-mABSA-T5 achieves the highest zero-shot accuracy (61.56%) across all task-specific models, and outperforms the Llama-3.1-8b-instant general-purpose model, despite using only 250M parameters and a tailored prompt. Though trailing larger general-purpose models like kimi-k2-instruct with 1 trillion parameters (Acc: 86%), Gen-mABSA-T5 remains highly reasonable, demonstrating that compact, prompt-engineered Flan-T5-base effectively handles Banglas morphological complexity in zero-shot ABSA, setting a new efficiency benchmark for low-resource languages.

5.4 Discussion and Observations

The experimental results reveal several key insights into the efficacy of zero-shot generative frameworks for multilingual ABSA, particularly in low-resource settings. First, consistently outperforms traditional task-specific transformer baselines in true zero-shot Bangla scenarios, demonstrating that prompt-engineered generative models can surpass language-specific discriminative models without any task-specific training. This advantage stems from Flan-T5's instruction-following capability and its exposure to NLI during pre-training, which allows ABSA to be naturally framed as a hypothesis–premise inference task.

Prompt ablation studies further confirm the critical role of prompt design. NLI-style prompts with concise and explicit instructions yield the highest average accuracy, outperforming direct prompts by 11–12 % and overly verbose prompts by 5–10 %. Adding domain context consistently degrades performance due to prompt dilution, reinforcing the principle that clarity and task focus are paramount in zero-shot settings.

Comparison with proprietary very large LLMs highlights a trade-off: GPT-OSS-20B (20B parameters) and Kimi-K2-Instruct (1T parameters), and Llama-8b-instant (8B parameters) achieve 61–85 % zero-shot accuracy on English SemEval tasks using conversational prompts, but require API access, higher latency, and non-deterministic outputs. Gen-mABSA-T5 (250M parameters), running locally on a single GPU, delivers 61-73 % accuracy with full reproducibility and no external dependencies, making it more suitable for resource-constrained research and deployment in low-resource regions.

6 Conclusion

Gen-mABSA-T5, a new multilingual zero-shot generative framework for Aspect-Based Sentiment Analysis (ABSA), demonstrates exceptional performance across English (SemEval 2014 Laptop and Restaurant) and Bangla (Large Bangla Corpus) datasets. It achieves a highest accuracy of 61.56% on the Large Bangla Corpus, improving to 86.32% with fine-tuning, while providing commendable accuracies of 73.50% (Laptop) and 73.56% (Restaurant) in English. Using prompt engineering and a generative transformer architecture, Gen-mABSA-T5 enables scalable, context-aware ABSA for low-resource languages like Bangla, addressing critical gaps in annotated resources and advancing multilingual sentiment analysis. Future work includes extending the framework to other underrepresented languages to further assess its generalizability

7 Limitations

The present study has several limitations that highlight areas for future improvement:

- The absence of a comprehensive and well-annotated Bangla ABSA data set limits the performance and generalization of the model due to data scarcity.
- The current prompts may not fully capture Bangla's complex morphological and syntactic characteristics, which can affect the precision of sentiment and aspect extraction.
- The model has not been evaluated on code-mixed or Romanized Bangla inputs (e.g., English and Bangla mixed sentences or Banglish), restricting its applicability in real-world multilingual communication contexts.

References

José I. Abreu Salas Agüero-Torales, Marvin M. and Antonio G. López-Herrera. 2021. Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing 107 (2021): 107373.*

Md Farhan Ishmam Navid Hasin Alvee Md Shahnewaz Siddique Md Azam Hossain Alam, Sadia and Abu Raihan Mostofa Kamal. 2024. Bnsentmix: A diverse bengali-english code-mixed dataset for sentiment analysis. *arXiv preprint arXiv:2408.08964 (2024).*

Rajib Chandra Dey and Orvila Sarker. 2019. Sentiment analysis on bengali text using lexicon based approach. *22nd International Conference on Computer and Information Technology.*

Zichen He Yaqin Xie Huan, Hai and Zelin Guo. 2022. A multi-task dual-encoder framework for aspect sentiment triplet extraction. *IEEE Access 10 (2022): 103187-103199.*

Mohammad Barkatullah Junayed Hossain, Sheikh Md Abdullah and Md Fahad Monir. 2024. Exploring the efficacy of bert in bengali nlp: A study on sentiment analysis and aspect detection. *IEEE EUROCON 2023-20th International Conference on Smart Technologies, pp. 54-59. IEEE, 2023.*

Xingyu Fu Guangluan Xu Yang Yang Jiuniu Wang Li Jin Qing Liu Li, Xinlong and Tianyuan Xiang. 2020. Enhancing bert representation with context-aware embedding for aspect-based sentiment analysis. *IEEE Access 8 (2020): 46868-46876.*

Ayesha Tasnim Mahfuz Ahmed Masum, Sheikh Junayed Ahmed and Md Saiful Islam. 2020. Banabsa: An aspect-based sentiment analysis dataset for bengali and its baseline evaluation. *arXiv preprint arXiv:2012.00288 (2020).*

Shamsul Arafin Mahtab, Nazmul Islam, and Md Mahfuzur Rahaman. 2018. Sentiment analysis on bangladesh cricket with support vector machine. pages 1–4. IEEE, 2018 international conference on Bangla speech and language processing (ICBSLP).

Tanjim Mahmud Mohammad Shahadat Hossain Md Akash Rahman, Manoara Begum and Karl Andersson. 2024. Analyzing sentiments in elearning: A comparative study of bangla and romanized bangla text using transformers. *IEEE Access (2024).*

Dhiman Sarma Rishita Chakma Md Morshedul Islam, GM Sakhawat Hossain and Md Mynoddin. 2023. Deep-absa: A multichannel deep learning framework for aspect-based bangla sentiment analysis. *World Conference on Communication Computing (WCONF), pp. 1-6. IEEE, 2023.*

Azizkhan F. Pathan and Chetana Prakash. 2022. Attention-based position-aware framework for aspect-based opinion mining using bidirectional long short-term memory. *Journal of King Saud University-Computer and Information Sciences 34, no. 10 (2022): 8716-8726.*

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Mumtahina Rahman Tuly Md Shihab Mahmud Rabeya, Mosa and Abdus Sattar. 2022. Sentiment analysis of bengali textual comments in field of sports using deep learning approach. *13th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-8. IEEE, 2022.*

Md Atikur Rahman and Emon Kumar Dey. Data 3, no. 2 (2018): 15. Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. *Applied Soft Computing 107 (2021): 107373.*

Md Atikur Rahman and Emon Kumar Dey. 2018. Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. *Data*, 3(2):15.

Shumaiya Akter Shammi, Sajal Das, Narayan Ranjan Chakraborty, Sumit Kumar Banshal, and Nishu Nath. 2023. A comprehensive roadmap on bangla text-based sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(4):1–29.

Maksuda Haider Sayma Md. Mohsin Kabir Shihab Ahmed, Moythry Manir Samia and M. F. Mridha. 2024. trf-bert: A transformative approach to aspect-based sentiment analysis in the bengali language. *Plos one, 19(9), p.e0308050.*

Lei Shu, Hu Xu, Bing Liu, and Jiahua Chen. 2022. Zero-shot aspect-based sentiment analysis. *arXiv preprint arXiv:2202.01924*.

Yang Deng Xin Li Yifei Yuan Lidong Bing Zhang, Wenxuan and Wai Lam. 2021. Aspect sentiment quad prediction as paraphrase generation. *arXiv preprint arXiv:2110.00796 (2021).*