ISSN 2007-9737

PromptMaster: Engineering Essentials and Basic NLP Tools

Shillpi Mishrra*, Anup Kumar Barman, Apurbalal Senapati

Central Institute of Technology, Kokrajhar, Assam,
India

shllpimishrra91@gmail.com, {ak.barman, a.senapati} @cit.ac.in

Abstract. Prompt Engineering (PE) and Large Language Models (LLM) are important developments in Natural Language Processing (NLP) research. This technique is key for crafting effective prompts that shape how the language model behaves. It is widely applied across various NLP tasks. Research has mainly focused on creating efficient prompts to boost performance in different applications, including chatbots, sentiment analysis, and text summarization. However, some fundamental questions remain unanswered, such as whether this work supports basic NLP or linguistic research in the context of PE and LLM. In traditional NLP, basic tools and techniques, such as part-of-speech taggers, named entity recognition, and morphological analyzers, are crucial for understanding language. Developing such tools remains a challenging issue, particularly for resource-scarce languages. In this paper, we are try to address this question. We have chosen Bengali as the language and have employed language models such as ChatGPT to tackle these challenges. For our experiments, we used publicly available datasets and the results were surprising when compared with the latest state-of-the-art models. We also identified the need to develop new prompts to fulfill basic requirements.

Keywords. Prompt engineering, natural language processing, Bengali, large language model.

1 Introduction

Natural language processing is part of artificial intelligence that concentrates on developing machines which can understand, interpret, and generate human language. Massive amounts of text or audio are subjected to operations and analysis with the help of machine learning, linguistics, and computer science concepts. Through the interface between human and machine languages, NLP seeks to make machines able to perform tasks, for example, sentiment analysis, translation, and summarization. A language can be characterized as a collection of a framework of guidelines that conveys information.

Natural Language Processing assists users who do not have the time to learn new languages, although not all users may be capable in machine-specific languages. It focuses on learning computers to understand words or phrases in human languages and sits at the interrelation of linguistics and artificial intelligence. It can be divided into two categories. One is natural language generation which enhances natural language understanding. Another is natural language understanding which designed to help users in their work and satisfy the desire to communicate with computers in natural language. Various concepts and methods that address the problem of natural language processing in the field of computer-mediated communication using natural language. Some of the NLP tasks that have been studied include morphological segmentation, which is the process of dissecting words into discrete morphemes that carry meaning. By determining its nouns, verbs, adjectives, etc.

Named Entity Recognition [29] analyzes a sentence and then connecting them to higher-order units with distinct grammatical meanings (verb groups, noun groups or phrases, etc.). Information is obtained using Named Entity Recognition, which identifies name entities and assigns them to various classes. Part-of-Speech Tagging [27] is a method for describing sentences that determines the part of speech for each word. The large language models [42] has increased dramatically at present

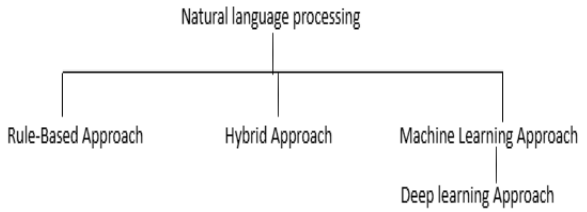


Fig. 1. NLP approach for different tasks

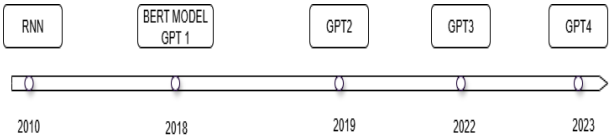


Fig. 2. Language Models timeline

time in a variety of contexts. These LLMs can carry out activities like language translation, content creation, conversational AI, etc. because they are made to process and produce languages that are equivalent to those of humans. As seen in Figure 1 [32], rule-based methods, machine learning algorithms, deep learning, hybrid approaches have been developed over time for handling limitations of NLP.

According to Hochreiter and Schmidhuber [18], the deep learning architectures (Short-Term Memory networks, and Gated Recurrent Units etc.) that have specifically introduced novel techniques across various NLP applications, as shown in Figure 3 . However, issues like lengthy training periods and the need for sizable datasets frequently make it difficult for RNN and LSTM models to capture important contextual information inside sequences.

Researchers used unsupervised neural networks to generate unique vector representations for words through embeddings in order to overcome the limitations in obtaining contextual information. In the beginning, methods such as Word2Vec, GloVe, and FastText ([32], [28]) produced word vectors, but they were unable to represent words without taking into consideration the context of the surrounding words or phrases.

NLP frameworks have developed to use contextual embeddings generated by extensive, pretrained language models in order to get around this. Since its introduction by [39], the Transformer model has developed as a key component of natural language processing, surpassing more established models such as CNNs and RNNs in tasks that require both language generation and understanding. For a variety of NLP tasks, such as text classification, language understanding, co-reference resolution, common-sense reasoning, and machine translation, its architecture is especially useful for pretraining on large text datasets, resulting in notable accuracy gains ([41], [25]). The capacity of NLP to comprehend context has greatly increased over time. More sophisticated models like GPT and BERT have significantly improved in overcoming pretraining difficulties for contextual comprehension, while earlier models like Word2Vec only moderately accurately captured word meanings.

An important turning point in the development of NLP was the switch from BERT to GPT-4. As shown in Figure 2, BERT's bidirectional approach transformed contextual comprehension, while the creativity and adaptability of GPT models such as GPT-1, GPT-3, and GPT-4 further advanced the field. This concept is enhancing AI's use in real-world scenarios.

While these advancements have driven remarkable progress in many widely studied languages, their impact on low-resource languages, such as Bengali, remains comparatively underexplored. Bengali (Bangla) is one of the Indo-Aryan languages spoken by over a million people. The basic Bangla alphabet consists of eleven vowels and forty-nine consonants, along with additional compound and numeric characters. These composite symbols are formed by combining consonants and vowels or consonants with one another. Bengali is an inflectional language that does not distinguish grammatical gender and relies heavily on prefixes and suffixes to express syntactic and semantic relationships.

While computational linguistics research for Bengali historically lagged behind high-resource languages like English, recent years have seen significant progress, with the development of

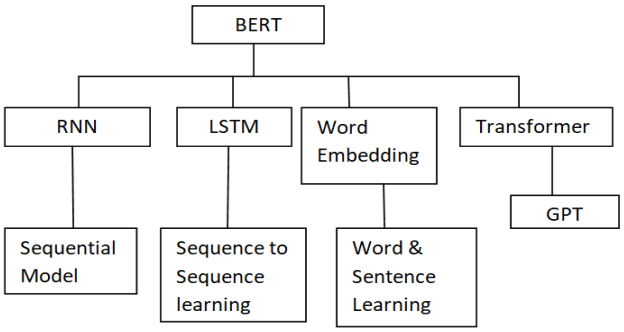


Fig. 3. Deep Tearning Approaches for NLP tasks: RNN to Transformer

various NLP tools, datasets, and models. As a result, Bengali is now considered a relatively well-resourced language in the South Asian context, with growing support for tasks such as part-of-speech tagging, named entity recognition and sentiment analysis.

However, despite these advancements, Bengali like many other languages still presents unique challenges for Large Language Models. Although LLMs exhibit extraordinary abilities in understanding and producing human language, their performance often skews toward high-resource languages, leaving gaps in accuracy, reasoning, and cultural nuance for languages such as Bengali. These challenges highlight the need for targeted evaluation and prompt-engineering strategies to fully harness LLMs for Bengali NLP tasks. Large Language Models continue to struggle with many of the world's languages, despite their remarkable progress in comprehending human language. This affects actual individuals; it's not just a technical glitch. Millions of speakers are deprived of the same level of machine translation, emotion analysis, and AI chatbots that English speakers take for granted when AI systems have trouble understanding these languages.Prompt engineering can help with it. We can contribute to closing this gap by properly crafting our prompts.

The problem, though, is how to create prompts that work for languages for which the AI initially has little training data. Teams from all around the world have been working on this problem and have developed innovative solutions that have a lot of potential.

Prompt Engineering [35] and Large Language Models have emerged as transformative approaches in Natural Language Processing, enabling significant advancements in tasks such as sentiment analysis, text summarization, and conversational agents.

However, most existing studies have primarily focused on high-resource languages and downstream tasks, leaving a critical gap in understanding whether LLMs and prompt-based techniques can effectively support foundational NLP tasks like Part-of-Speech tagging, Named Entity Recognition, and Morphological Analysis, especially in low-resource language contexts.

This paper addresses the underexplored question: Can prompt engineering with LLMs replace traditional tools for foundational NLP tasks in low-resource languages like Bengali? Our work focuses on the Bengali language, which lacks mature NLP resources, and investigates how well general-purpose LLMs like GPT model perform on core linguistic tasks.

To this end, we:
(A) Design and evaluate tailored prompts for basic NLP tasks in Bengali.
(B) Evaluate the performance of LLM-generated outputs with existing state-of-the-art models using publicly available Bengali datasets.
(C) We analyze existing limitations and introduce prompting strategies like zero-shot and chain-of-thought to enhance efficiency in core NLP tasks.

2 Background

2.1 Basic NLP tools

Computational analysis and comprehension of human language depend on Natural Language Processing techniques. In order to process languages, including Bengali, three essential NLP techniques are required: Named Entity Recognition, Morphological Analysis, and Parts-of-Speech tagging. A explanation of each tool is provided below, along with Bengali examples:

2.1.1 Named Entity Recognition

The recognizing various informational segments mentioned in a text and then categorizing them into pre-established groups is known as named entity recognition. In a text, NER recognizes proper names, locations, organizations, and other named entities. It classifies words into predetermined categories, including names of people, places, organizations, and dates. The lack of capitalization hints and spelling changes in Bengali make NER especially difficult.

Let us consider an example.

Sentence: "রবীন্দ্রনাথ ঠাকুর শান্তিনিকেতনে বসবাস করতেন।
(Translation: "Rabindranath Tagore lived in Santiniketan.")

NER Output is the following:
রবীন্দ্রনাথ ঠাকুর (Rabindranath Tagore) → PERSON,
শান্তিনিকেতন (Santiniketan) → LOCATION.

2.1.2 Part-of-Speech Tagging

These types of tasks assign an appropriate grammatical category to each and every word in a text. POS tags are also referred to as morphological classes and lexical tags. They all play a crucial role in language processing by representing the word and its environment [21, 31]. This technique is widely utilized in applications such as information retrieval, Text-to-Speech systems, shallow parsing, and information extraction.

Moreover, POS tagging helps linguistic research on corpora. It serves as a foundational step for more advanced NLP tasks like semantic analysis, parsing and machine translation [16]. Therefore, for complex NLP applications in Bangali and other languages, POS tagging serves as an essential resource.

Let us consider an example.

Sentence: "সে সুন্দর ফুল দেখলো।" (Translation: "He/She saw a beautiful flower.")

POS Tagging Output is the following:
সে (Se) → PRON (Pronoun),
সুন্দর (Sundor) → ADJ (Adjective),
ফুল (Phul) → NOUN (Noun),
দেখলো (Dekhlō) → VERB (Verb - Past Tense).

2.1.3 Morphological Analysis

Morphology is the approach which tells how words are formed by concentrating on their inherent structure. The two classes of morphology are derivational morphology and inflectional morphology. When a word stem and a morpheme are united in inflectional morphology, the resulting word has the same class as the word stem; in derivational morphology, however, the resulting word has a different class than the word stem [24].

Let us consider an example.

Word: "খেলছিলাম" (khelchilam - "was playing")

Morphological Analysis Output is as follows:
Root: খেলা (khela - "play"),
Suffixes: ছিল (chil - past tense marker) + আম (am - first-person singular marker).

2.2 Large Language Model

An artificial intelligence model builds on extensive data to understand and produce human-like language is known as a Large Language Model. Similar models, including GPT series, learn to understand and predict language patterns through deep learning methods, especially transformer architectures. LLMs are capable of effectively doing many things, answering a question, summarizing a text, translating a language, writing a creative paragraph, or some code.These models use billions of parameters to understand contextual meanings, subtle nuances, and word relationships.

This helps them create coherent and contextually appropriate responses. Over time, as they become more capable, LLMs are applied in various areas, changing fields like customer service, education, and content creation. While these systems assist us in many ways, they also raise ethical concerns related to bias, misinformation, and responsible use. This means we need careful development and oversight in how we use them.

These powerful language models, which have large parameter sizes, are specifically trained to learn a great deal [6, 23, 42]. The architecture behind many LLMs includes models like GPT-3 [14], InstructGPT [30], and GPT-4 [1].

These models use self-attention as the main method for language modeling. Transformers [2] have changed natural language processing by

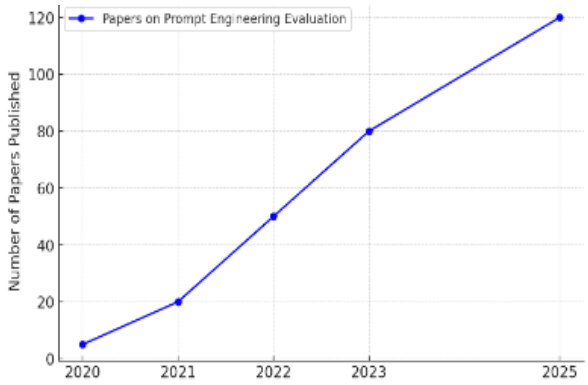


Fig. 4. Trends in LLMs using prompt engineering evaluation papers from 2020 to February 2025

effectively handling sequential data, enabling parallel processing, and capturing long-range relationships within text.

In-context learning [5] is one of the most important abilities of LLMs. It allows a model to create text step-by-step based on earlier prompts or context that is improving coherence and relevance. This feature is especially useful for interactive tasks and conversations.

Reinforcement learning with human input [7, 43] is crucial for enhancing LLMs. The model has been trained using data from October 2023 to learn how to improve itself based on human responses received as rewards. It adjusts one layer at a time, following a natural learning cycle where mistakes become valuable.

Prompt engineering is basically instructions for Large Language Models that organize and improve interactions with LLMs. This is done through effective input prompts that guide the model's responses. LLMs like GPT model use in-context learning for further text generation. This makes the training prompt crucial for their performance [5]. By designing specific prompts, users can shape LLM behavior without needing to fine-tune the model. As a result, this method serves as a powerful tool for many applications, including text generation, question answering, and even automated reasoning [34]. In addition, prompt engineering plays an important role in reducing bias and increasing task-level accuracy through their ability to fine-tune performance through few- and zero-shot learning, which requires fewer examples to produce high-quality outputs [15].

2.3 Prompt Engineering

Here we provide a concise overview of different prompting techniques and their growing benefits towards improvement of the model performance. Most of these prompting strategies were evaluated on at least two main techniques — zero-shot and few-shot.

Certain techniques specifically lend themselves to zero-shot versus few-shot applications (or vice versa), with no other reasonable options. According to [33, 35], the language model in the zero-shot technique is given no training data and must rely only on its prior knowledge to finish tasks when given explicit instructions.

As mentioned by [5], the few-shot setting aims to enhance the model's understanding by offering it task-specific prompts along with a limited number of training instances. Although studies have demonstrated that few-shot prompting frequently results in better performance, it also presents difficulties, such as possible biases toward the chosen few-shot examples, requiring thorough data preparation.

2.3.1 Zero-shot Prompting Technique

Zero-shot prompting represents a transformative approach to utilizing large LLMs. Rather than requiring extensive training data, this technique relies on well-designed prompts to direct the model toward performing new tasks [33].

This method involves providing the model a task description in the prompt, but it does not include labeled data for training on specific input-output relationships. Instead, it utilizes its pre-trained knowledge to generate predictions specifically based on the given prompt, enabling it to handle new tasks effectively.

2.3.2 Few-shot Prompting Technique

Zero-shot prompting gives models no examples. But few-shot prompting gives models a small number of input-output examples to help them comprehend a particular task [5]. When it comes to challenging tasks, having even a limited quantity of high-quality examples is valuable for model performance in comparison to the absence of demonstrations. To accommodate the examples, this method necessitates extra tokens, which may not be feasible for larger text inputs.

Furthermore, model behavior is greatly influenced by the selection and organization of prompt instances, and biases—like a predilection for frequent words—may still affect outcomes. Effective prompt engineering is crucial to maximize results and minimize unintentional biases, even while few-shot prompting enhances performance on challenging tasks, especially for big pre-trained models like GPT-3.

2.3.3 Chain-Of-Thought Technique

Wei et al. [40] focus on the idea that people inherently divide difficult problems into smaller and easier-to-manage subproblems before arriving up with a final solution in their prompting approach. In a similar way, the authors investigate how producing a organized series of intermediate reasoning steps that can greatly enhance LLMs' reasoning ability. Their results show a significant improvement over simple prompting, with the largest performance difference for mathematical problem-solving tests being about 38 percentage and for commonsense reasoning tasks being about 26 percentage.

3 Early Work

An open-source finite-state morphological analyzer was developed for Bengali language that is described in the paper [13]. The authors create a system using the open-source finite-state toolkit lttoolbox in order to overcome these difficulties. They then test the system's ability to accurately interpret Bengali morphology in order to assess its effectiveness. Lastly, they offer possible upgrades to improve the analyzer's coverage and efficacy. In this paper, the problem is that the analyzer requires a lot of work and is still in the early stages. Its coverage is restricted, requiring the human tagging of many additional phrases. It has trouble with multi-word verbs, particularly when the negative particles are positioned differently. The complex structure of Bengali verb negation makes it difficult to distinguish the negative forms of verbs.

A better, more effective method is required because trying to handle this with layered paradigms slows down the analyzer [38].

To enhance Named Entity Recognition (NER) in Bengali, the researchers [37] tested with several characteristics, including POS tags, word suffixes and word embeddings. After optimizing the BanglaBERT (large) model for this task, they found that deep learning models are capable of capturing intricate patterns that are difficult for humans to recognize. Their optimized model produced satisfactory performance, reaching an F1 score of 0.79. The significance of Bengali Complex Named Entity Recognition (CNER) is emphasized in the study, particularly when synthetic datasets are used. All things considered, it shows how well deep learning techniques like BanglaBERT work for Bengali NER problems. According to the paper, deep learning models outperformed conventional techniques like CRF on Bengali CNER. It also implies that in order to attain high accuracy, deep learning techniques need additional resources and fine-tuning. Furthermore, because the data may not accurately reflect authentic language use, relying solely on a synthetic dataset may restrict real-world applicability. Lastly, the field's low representation in NLP research suggests that there aren't many tools and resources available for Bengali.

Bengali NLP research primarily concentrates on specific areas such as optical character recognition, sentiment analysis, speech recognition and text summarization. Nonetheless, comprehensive resources that evaluate the latest BNLP tools and techniques are noticeably lacking [36]. This paper categorized into 11 groups such as POS tagging, Information Extraction, Machine Translation, Sentiment Analysis and Speech Processing. Using a variety of datasets, they examine traditional

approaches including machine learning and deep learning methods. While discussing the constraints and new developments in BNLP.The field of BNLP offers several advantages such as its growing importance due to the widespread use of Bangla and making it increasingly relevant on a global scale. The paper provides a thorough review that covering various preprocessing techniques and classical as well as machine learning methods.

It is offering valuable insights into commonly used approaches and identifying new areas that remain unexplored. It also discusses the limitations in the development of BNLP systems and presents future research possibilities. Furthermore, it highlights the complexities of BNLP and the necessity of understanding language characteristics in modern challenges. On the downside, during the development of BNLP systems the paper acknowledges the significant limitation faced. The complexities involved along with some areas that still need more exploration. It suggests that BNLP requires focused research and improvement to fully unlock its potential.

The authors [8] define how a morphological analyzer can be more successful by combining contextual information such as surrounding words or syntactic context. This process also includes character-level properties like individual letters or character sequences in the Bengali language. This method helps the system's understanding and process out-of-vocabulary words. These are words that it did not encounter during training. It also increases accuracy when handling words with several morphological forms (e.g., distinct derivations or inflections). The program can produce more accurate predictions by utilizing both comprehensive character data and more extensive sentence context. Lower performance and irregularities in morphological feature agreement may result from the existing system's inadequate integration of Bengali-specific linguistic features.

The paper [20] aims to develop and compare supervised traditional approaches for Bangla part-of-speech tagging. Utilizing a large dataset, the research evaluates tagging accuracy to identify the most effective method for enhancing POS tagging performance in BNLP applications. Among the models tested, the Long Short-Term Memory network is achieved the highest accuracy is 95.60 percentage. However, the study is limited in scope, as it evaluates individual tagging techniques in isolation, without investigating hybrid or ensemble models. Additionally, it focuses exclusively on POS tagging in the Bangla language, without extending to multilingual contexts. Future research could address the challenge of improving tagging accuracy for rare or ambiguous words.

The paper [38] aims to conduct a morphological comparison between English and Bangla from synchronic and grammatical perspectives. The primary objective is to assist second language learners in understanding the structural differences between the two languages and overcoming challenges in second language acquisition. The study emphasizes that learning English as an L2 is essential yet often difficult due to factors such as mother tongue interference, age, and motivation. Through its comparative analysis, the research provides valuable insights into addressing these challenges. By examining English and Bangla across multiple linguistic dimensions, the study contributes to improved comprehension and acquisition for L2 learners. However, the study is constrained by its small scale, exploratory nature, limited financial and temporal resources, and a lack of comprehensive study materials—particularly concerning Bangla morphology and syntax—highlighting the need for more extensive and rigorous future research.

The paper [22] aims to develop an efficient and accurate Part of Speech (POS) tagger for the Bengali language using deep learning techniques that leveraging morphological and contextual information to address linguistic ambiguities and improve tagging accuracy for natural language processing applications. This paper uses Deep Belief Network (deep learning-based model) and the model gives 93.33 Percentage accuracy. The study is limited by data sparsity, with some POS tag classes underrepresented in the corpus, and by class imbalance, leading to potential bias toward frequent tags. Additionally, the lack of availability of comparable corpora prevented direct comparison with previous POS tagging methods.

The paper [19] aims to develop an automated Part-of-Speech tagging system for low-resource

Bangla that utilizing a suffix-based approach combined with a custom stemming technique. It uses an extensive vocabulary database, and verb pattern datasets. This model which is based on rules that aims to enhance the precision and speed of tagging in Natural Language Processing tasks. It achieves an accuracy rate that is 93.7 percent. However, the system is limited in scope, supporting only eight fundamental POS tags without addressing subcategories or punctuation—both of which are critical for more nuanced NLP tasks. Additionally, the approach does not incorporate probabilistic models, which could potentially improve overall tagging accuracy.

The paper [9] focuses on developing a Bengali Noun Morphological Analyzer using a linguistic approach grounded in finite-state transducer grammar. The objective is to analyze nominal suffixes and enhance the morphological analysis of Bengali nouns by expanding linguistic resources. The analyzer is based on a set of FST grammar rules and achieved an accuracy of 44 percentage. However, the tool has several limitations. It relies on a fixed set of transducer grammar rules and is restricted to noun morphology, without extending to other parts of speech. Additionally, it lacks the ability to analyze named entities or complex noun group structures.

The system does not yet support broader basic NLP tasks for Bengali that indicating significant scope for future development. The paper [10] is to develop an unsupervised morphological analysis algorithm for Bengali that effectively segments words into prefixes, suffixes, and stems without prior linguistic knowledge. The goal is to for highly inflectional languages such as Bengali. The approach focuses on morpheme induction and word segmentation, achieving an accuracy of 64.62 percentage. However, the study has several limitations. It struggles to handle highly irregular word forms and does not incorporate semantic information, which leads to attachment errors. Furthermore, the approach is constrained by the limited availability of annotated corpora for Bengali, which restricts comprehensive empirical evaluation and hampers the scalability and robustness of the system.

The paper [26] is to develop and evaluate a Grammatical Error Explanation system for Bengali that not only corrects grammatical errors but also provides meaningful natural language explanations. To support this goal, the researchers introduced a multi-domain dataset and benchmarked various large language models to enhance language learning in low-resource settings. The system achieved an accuracy of 74 percentage with GPT model and other LLMs forming the backbone of the evaluation. Despite this progress, the study found that these models struggle with nuanced Bengali grammatical errors, such as Guruchondali dosh, incorrect case markers — particularly in short sentences containing multiple errors. This study [3] focuses on developing and evaluating deep learning models for Named Entity Recognition in Bangla. It uses BERT-based contextual embeddings. To tackle the common challenges in low-resource languages like Bengali, the study employs a modified cost-sensitive loss function.

It also experiments with Conditional Random Fields and Focal Loss. The proposed approach combines BERT, BiLSTM, CRF, and class weighting, achieving an accuracy of 73 percent. While the cost-sensitive learning strategy improves F1 Macro scores and helps with class imbalance, the model was mainly evaluated on limited datasets and task scopes. Therefore, further validation with more diverse datasets and languages is needed to prove its broader applicability and generalizability.

The goal of this paper [17] is to create a Bangla Named Entity Recognition (B-NER) dataset that includes eight different entity types. This addresses the limitations of earlier systems, which were limited to just three types. The dataset is manually labeled using real-world data and aims to support more research in Bangla NER. The accuracy of Bidirectional Long Short-Term Memory model is 86 percent. Despite its comprehensiveness, the B-NER dataset faces challenges related to class imbalance and limited size which may hinder the generalizability of trained models. Future work is directed toward expanding and balancing the dataset to improve its utility for robust NER system development.

4 Experiment

This research aims to assess whether prompt engineering with Large Language Models can effectively replace traditional tools for foundational NLP tasks in low-resource languages, with a particular focus on Bengali. Although Bengali has seen progress in NLP research, it is an ideal test case for evaluating the adaptability of general-purpose LLMs such as the GPT family.

In our experiments, we designed and evaluated customized Bengali prompts for basic NLP tasks. We then compared the performance of LLM-generated outputs with that of existing state-of-the-art models using publicly available Bengali datasets. To address the inherent limitations observed in baseline performance, we further incorporated zero-shot and chain-of-thought prompting strategies, to strengthen the model's reasoning ability and boost its effectiveness in low-resource linguistic contexts.

4.1 Dataset

Bengali is one of the prominent languages of the family of Eastern Indo-Aryan languages. We used the BanglaBlend dataset [4] for our experiment.

This dataset contains basically 7,350 annotated Bangla sentences and categorized sentences between the Common (Cholito) and Saint (Sadhu) forms.

The dataset provides information for developing language technologies like POS taggers and NER for studying Bengali language structures with computational methods.

The structure of the dataset is defined in Table 1 as well as the structure of the sentence. Each sentence is categorized with sadhu and cholito classification.

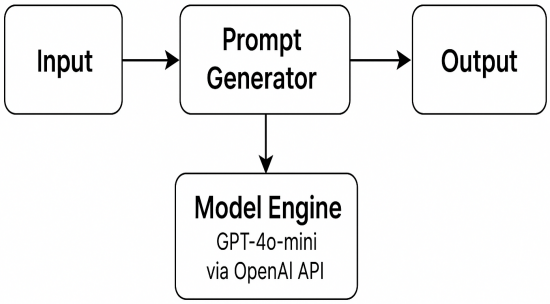


Fig. 5. System architecture of our work

4.2 Experimental Setup

The experimental setup is designed to test how well large language models, particularly GPT-4o-mini, can perform basic NLP tasks for the Bengali language using prompt engineering techniques. The model architecture, shown in Figure 5, has four main components. T

he components are input module, prompt generator, model engine, and output handler. The input module takes raw Bengali sentences from the BanglaBlend dataset. This dataset contains annotated samples in both Sadhu (formal) and Cholito (colloquial) styles. The prompt generator receives the inputted text. This module converts them into structured prompts for POS tagging, Named Entity Recognition, and Morphological Analysis.

We use two prompting strategies: Zero-Shot and Chain-of-Thought. These strategies guide the model's reasoning. The prompts will be processed by the model engine using OpenAI's GPT-4o-mini model.

The LangChain framework is used to communicate with the language model. For evaluation, the output handler module captures and interprets the model's responses. It extracts the structured results. In low-resource language contexts, using prompt-based NLP solutions, the modular setup provides a flexible and scalable way to experiment. The system mainly depends on OpenAI's GPT-4o-mini language model, which is accessed through the OpenAI API.

Table 1. Examples of the BanglaBlend Dataset. Dataset Attributes with Examples of Sadhu and Cholito Sentence Classification

Attribute	Description
Sentence:	1. ফুটবলে এটিকে বলা হয় কিলার বল বা ফাইনাল বল 2. সঙ্গে বাদ্য থাকে তাহা হইলে বর কনে দেখিবার নিমিত্ত পাল্কীর ভিতর দৃষ্টিপাত করে
Label:	Each sentence is classified as Sadhu (সাধু) and Cholito (চলিত) for example: 1. ফুটবলে এটিকে বলা হয় কিলার বল বা ফাইনাল বল → Cholito (চলিত) 2. সঙ্গে বাদ্য থাকে তাহা হইলে বর কনে দেখিবার নিমিত্ত পাল্কীর ভিতর দৃষ্টিপাত করে → Sadhu (সাধু)

By using the LangChain framework, the prompt construction and response handling would be managed in a modular way. This framework helps to make the interaction between the language models and the external data sources. The system is implemented with the help of Python 3.10 which is supported by several libraries.

These included openai for API communication, langchain for managing prompts, dotenv for secure handling of API keys through a .env file, and standard libraries like json and pandas for data formatting and analysis. We set the temperature in the experimental setup to 0.7, keeping other decoding parameters, such as top-p, at their default values. These settings maintained consistency across experiments while allowing for detailed control over the model's creativity and response diversity. All experiments took place in Google Colab, providing a cloud-based and GPU-enabled environment that supported rapid prototyping, reproducibility, and effective resource use.

The prompt:

আপনি একজন ভাষাবিজ্ঞান বিশেষজ্ঞ। নিচের পাঠ্যটির Part-of-Speech (POS) ট্যাগিং করুন:

সকলেরই অনাবৃত দেহ সকলের সেই অনাবৃত বক্ষে আরশির ধুকধুকি চন্দ্রকিরণে এক একবার জ্বলিয়া উঠিতেছ

Response:

POS Tagging (Word/POS):

১. **সকলেরই** - Pronoun
২. **অনাবৃত** - Adjective
৩. **দেহ** - Noun
৪. **সকলের** - Pronoun
৫. **সেই** - Pronoun
৬. **অনাবৃত** - Adjective
৭. **বক্ষে** - Noun
৮. **আরশির** - Noun
৯. **ধুকধুকি** - Noun
১০. **চন্দ্রকিরণে** - Noun
১১. **এক** - Numeral
১২. **একবার** - Noun
১৩. **জ্বলিয়া** - Verb
১৪. **উঠিতেছ** - Verb

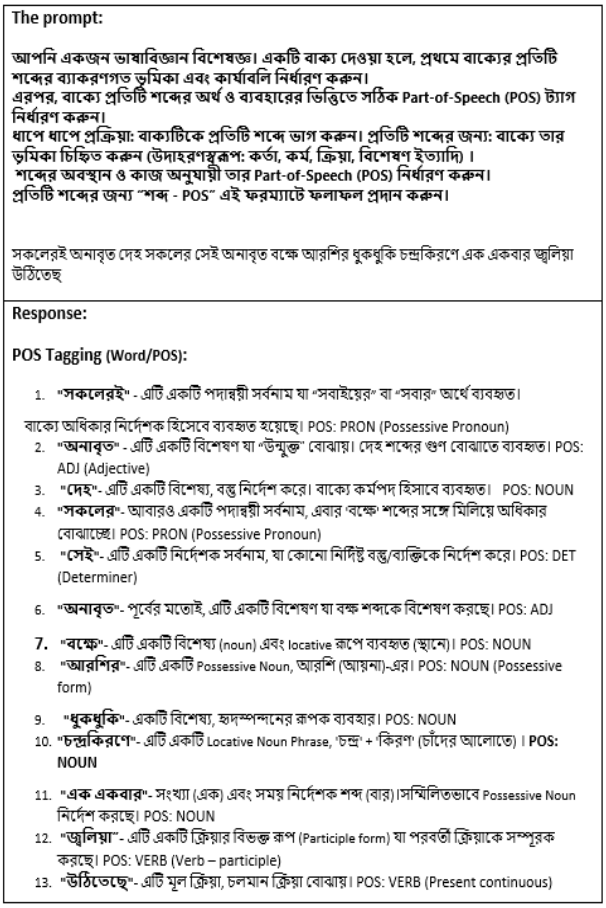
Fig. 6. POS tagged analysis with the Zero-shot prompt

4.3 Evaluation Metrics

We use the bengali prompting techniques for performing basic NLP tasks on Bengali language. Precision, Recall, and F1-Score help us to examine the model's ability to perform basic tasks like parts-of-speech tagging, morphological analysis, and named entity recognition. They perform these tasks by generating acceptable and relevant outputs.

1. Precision is the proportion of actually expected positive calls to all those being classified as positive predicted. This metric offers the following query: What percentage of the objects to which the model has applied a certain tag are in fact correct?

2. The percentage of positive examples in the data that were accurately predicted is known



The prompt:

আপনি একজন ভাষাবিজ্ঞান বিশেষজ্ঞ। একটি বাক্য দেওয়া হলে, প্রথমে বাক্যের প্রতিটি শব্দের ব্যাকরণগত ভূমিকা এবং কার্যাবলি নির্ধারণ করুন।
এরপর, বাক্যে প্রতিটি শব্দের অর্থ ও ব্যবহারের ভিত্তিতে সঠিক Part-of-Speech (POS) ট্যাগ নির্ধারণ করুন।
ধাপে ধাপে প্রক্রিয়া: বাক্যটিকে প্রতিটি শব্দে ভাগ করুন। প্রতিটি শব্দের জন্য: বাক্যে তার ভূমিকা চিহ্নিত করুন (উদাহরণস্বরূপ: কর্তা, কর্ম, ক্রিয়া, বিশেষণ ইত্যাদি)।
শব্দের অবস্থান ও কাজ অনুযায়ী তার Part-of-Speech (POS) নির্ধারণ করুন।
প্রতিটি শব্দের জন্য "শব্দ - POS" এই ফরম্যাটে ফলাফল প্রদান করুন।

সকলেরই অনাবৃত দেহ সকলের সেই অনাবৃত বক্ষে আরশির ধুকধুকি চন্দ্রকিরণে এক একবার জ্বলিয়া উঠিতেছ

Response:

POS Tagging (Word/POS):

1. "**সকলেরই**" - এটি একটি পদান্বয়ী সর্বনাম যা "সবাইয়ের" বা "সবার" অর্থে ব্যবহৃত।

বাক্যে অধিকার নির্দেশক হিসেবে ব্যবহৃত হয়েছে। POS: PRON (Possessive Pronoun)

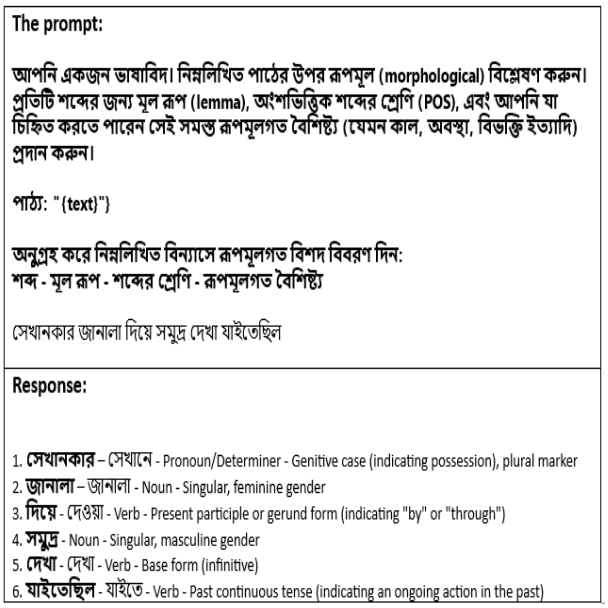
2. "**অনাবৃত**" - এটি একটি বিশেষণ যা "উন্মুক্ত" বোঝায়। দেহ শব্দের গুণ বোঝাতে ব্যবহৃত। POS: ADJ (Adjective)
3. "**দেহ**"- এটি একটি বিশেষ্য, বস্তু নির্দেশ করে। বাক্যে কর্মপদ হিসাবে ব্যবহৃত। POS: NOUN
4. "**সকলের**"- আবারও একটি পদান্বয়ী সর্বনাম, এবার 'বক্ষে' শব্দের সঙ্গে মিলিয়ে অধিকার বোঝাচ্ছে। POS: PRON (Possessive Pronoun)
5. "**সেই**"- এটি একটি নির্দেশক সর্বনাম, যা কোনো নির্দিষ্ট বস্তু/ব্যক্তিকে নির্দেশ করে। POS: DET (Determiner)
6. "**অনাবৃত**"- পূর্বের মতোই, এটি একটি বিশেষণ যা বক্ষ শব্দকে বিশেষণ করছে। POS: ADJ
7. "**বক্ষে**"- এটি একটি বিশেষ্য (noun) এবং locative রূপে ব্যবহৃত (স্থানে)। POS: NOUN
8. "**আরশির**"- এটি একটি Possessive Noun, আরশি (আয়না)-এর। POS: NOUN (Possessive form)
9. "**ধুকধুকি**"- একটি বিশেষ্য, হৃদস্পন্দনের রূপক ব্যবহার। POS: NOUN
10. "**চন্দ্রকিরণে**"- এটি একটি Locative Noun Phrase, 'চন্দ্র' + 'কিরণ' (চাঁদের আলোতে)। **POS: NOUN**
11. "**এক একবার**"- সংখ্যা (এক) এবং সময় নির্দেশক শব্দ (বার)।সম্মিলিতভাবে Possessive Noun নির্দেশ করছে। POS: NOUN
12. "**জ্বলিয়া**"- এটি একটি ক্রিয়ার বিভক্ত রূপ (Participle form) যা পরবর্তী ক্রিয়াকে সম্পূরক করছে। POS: VERB (Verb – participle)
13. "**উঠিতেছ**"- এটি মূল ক্রিয়া, চলমান ক্রিয়া বোঝায়। POS: VERB (Present continuous)

Fig. 7. POS tagged analysis with the Chain-of-Thought prompt

as recall. In simple terms, it asks: How many items did the model identify correctly if we take into consideration all the real items that should have been assigned a particular label? Even if some incorrect examples are included, a high recall shows that the model is very responsive and receives the majority of the appropriate instances.

3. A key measure that illustrates how effectively the compromises balance out is the F1-score, which is a harmonic mean of recall (Rn) and precision (Pn). It is calculated by this formula:

$$F_1 - score = 2 \times \frac{Pn \times Rn}{Pn + Rn}. \qquad (1)$$



The prompt:

আপনি একজন ভাষাবিদ। নিম্নলিখিত পাঠের উপর রূপমূল (morphological) বিশ্লেষণ করুন। প্রতিটি শব্দের জন্য মূল রূপ (lemma), অংশভিত্তিক শব্দের শ্রেণি (POS), এবং আপনি যা চিহ্নিত করতে পারেন সেই সমস্ত রূপমূলগত বৈশিষ্ট্য (যেমন কাল, অবস্থা, বিভক্তি ইত্যাদি) প্রদান করুন।

পাঠ্য: "{text}"}

অনুগ্রহ করে নিম্নলিখিত বিন্যাসে রূপমূলগত বিশদ বিবরণ দিন:
শব্দ - মূল রূপ - শব্দের শ্রেণি - রূপমূলগত বৈশিষ্ট্য

সেখানকার জানালা দিয়ে সমুদ্র দেখা যাইতেছিল

Response:

1. **সেখানকার** – সেখানে - Pronoun/Determiner - Genitive case (indicating possession), plural marker
2. **জানালা** – জানালা - Noun - Singular, feminine gender
3. **দিয়ে** - দেওয়া - Verb - Present participle or gerund form (indicating "by" or "through")
4. **সমুদ্র** - Noun - Singular, masculine gender
5. **দেখা** - দেখা - Verb - Base form (infinitive)
6. **যাইতেছিল** - যাইতে - Verb - Past continuous tense (indicating an ongoing action in the past)

Fig. 8. Morphology Analysis with the Zero-shot prompt

These metrics were calculated separately for each task, including POS tagging, morphological analysis, and NER, using both prompting strategies. The results were compared with those of traditional machine learning models mentioned in previous studies. This comparison helps us assess not only the overall performance of GPT-4o-mini but also how it measures up against domain-specific models developed through conventional approaches.

5 Result

In this study, we describe how effectively LLMs, specifically the GPT-4o model, using prompt techniques can perform on basic natural language processing tasks such as NER, POS tagging, and Morphological Analysis for the Bengali language. To help the LLM in completing tasks, we apply two Bengali prompting techniques that are Zero-Shot Prompting and CoT Prompting.

We use the publicly available BanglaBlend dataset [4] which contains a huge range of Bengali texts. The dataset is categorized into Sadhu Bhasha (formal) and Cholit Bhasha (colloquial).

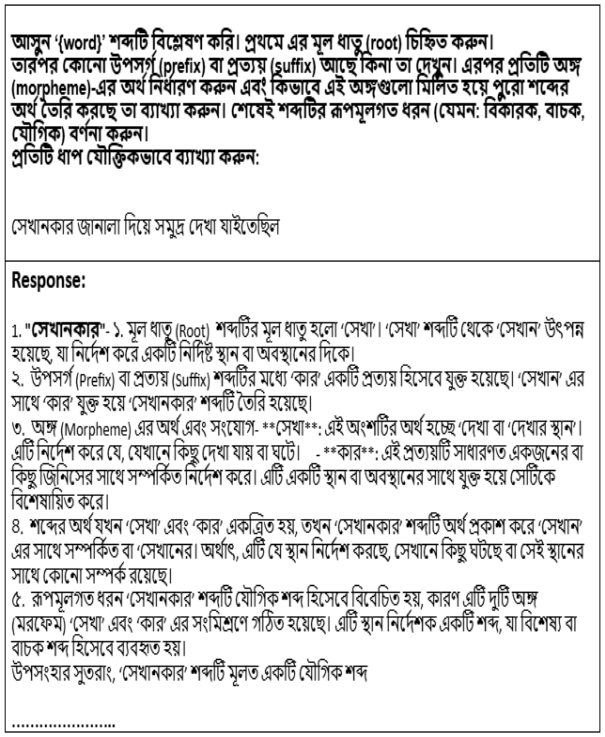


Fig. 9. Morphology Analysis with the Chain-of-thought prompt

5.1 POS Tagging

To analyze parts of speech tagging on the GPT-4o-mini model, we work on two prompting techniques that are zero shot and chain-of-thought prompt. This helps us to understand how well the model identifies grammatical categories of words with different levels of prompt guidance.

5.1.1 Zero-Shot Prompt

This prompt is a direct instruction which is being sent to the model without examples or intermediate reasoning for analysis pos-tagging. In Figure 6, we define Bengali the zero-shot prompt for POS tagging analysis.

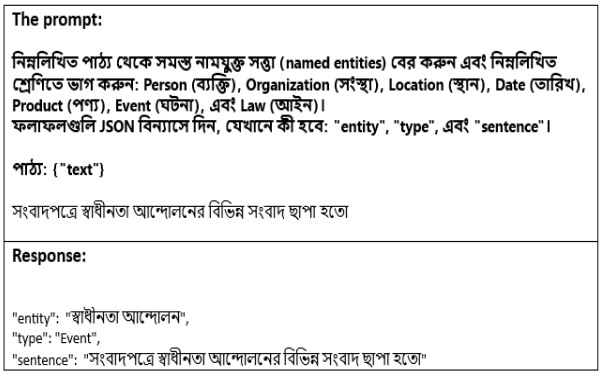


Fig. 10. NER Analysis with the Zero-shot prompt

5.1.2 Chain-of-Thought

We instruct the model using Chain-of-Thought prompting to perform the step-by-step reasoning. In Figure 7, the Bengali chain-of-thought prompt is used for POS tagging analysis that has both the input and the corresponding output.

5.1.3 Morphological Analysis

An important basic task of natural language processing is morphological analysis, which aims to decompose words into their fundamental meaningful components that are called morphemes (roots, prefixes, suffixes). This helps in understanding the word structure, meaning, and its grammatical function.

In this study, we utilize GPT-4o-mini to perform morphological analysis developed by OpenAI.

We explore the model's performance using two different prompting strategies:
(a) Zero-shot prompting,
(b) Chain-of-Thought prompting.

The goal is to compare the effectiveness of each technique in helping the GPT-4o-mini model accurately perform morphological analysis.

5.1.4 Zero-Shot Prompt

In this setting, we provided a direct instruction to the model without examples or intermediate reasoning. Figure 8 presents the zero-shot prompt for morphological analysis, detailing the input and its corresponding output.

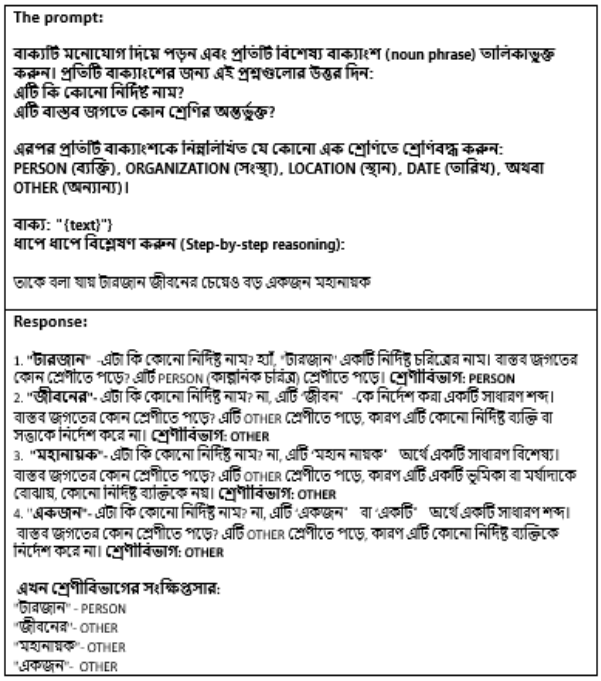


Fig. 11. NER Analysis with the Chain-of-Thought prompt

5.1.5 Chain-of-Thought Prompt

we asked the model to perform a step-by-step reasoning using Chain-of-Thought prompting. Figure 9 presents the chain-of-thought prompt for morphological analysis, detailing the input and its corresponding output.

5.2 NER Analysis

A basic task in Natural Language Processing is Named Entity Recognition. This involves recognizing and categorizing entities in the text into established categories like personal names, organizations, geographical locations, dates, etc.

Traditionally tackled using machine learning or rule-based models, modern transformer-based models like GPT-4o-mini developed by OpenAI offer a new way to approach NER using prompting techniques.

We explore the model's performance using two different prompting strategies:
(a) Zero-shot prompting,
(b) Chain-of-thought prompting.

5.2.1 Zero-Shot Prompt

In this setting, we provided a direct instruction to the model without examples or intermediate reasoning. Figure 10 presents the zero-shot prompt for NER analysis, detailing the input and its corresponding output.

5.2.2 Chain-of-Thought Prompt

We instructed the model to engage in step-by-step reasoning using chain-of-Thought prompting. Figure 11 presents the chain-of-thought prompt for morphological analysis, detailing the input and its corresponding output.

Although all prompts in this study were originally designed in Bengali, we also translated and tested them in English. The results remained consistent across both languages.

In Table 2, our results using Bengali prompting techniques have been compared with the existing state-of-the-art where the results for our system are displayed in the first two rows, while the results for one of the most advanced techniques currently in use are displayed in the third row. The evaluation covers three foundational natural language processing tasks: Part-of-Speech Tagging, Morphological Analysis, and Named Entity Recognition for the Bengali language. Chain-of-Thought prompting consistently outperforms Zero-Shot prompting across all three tasks that is showed in the result section and indicating that step-by-step reasoning helps the model generate more accurate and structured outputs. For instance, in POS tagging and morphological analysis, CoT prompts yield significantly higher F1-scores, demonstrating their effectiveness in handling grammatical and inflectional complexities of Bengali.

Compared to traditional models, GPT-4o-mini with CoT prompting performs competitively or even better in several cases. Specifically, while SVM and LSTM models have shown strong performance in earlier benchmarks, the language model's ability to generalize through in-context learning and rich language understanding allows it to surpass these approaches in certain scenarios—particularly in POS tagging and morphological segmentation.

Table 2. Comparison of NLP Task Performance using Bengali Zero-Shot prompt and Chain-of-Thought Prompt with GPT-4o model versus existing state-of-the-art

NLP Tools	Methodologies	Model	Precision	Recall	F1-Score
POS-tagging	Bengali prompt using Zero-Shot	GPT-4o-mini	85.0	54.0	66.04
	Bengali prompt using Chain-of-Thought	GPT-4o-mini	91.4	74.5	82.08
	Deep learning [20]	LSTM	70.0	73.0	71.0
Morphology	Bengali prompt using Zero-Shot	GPT-4o-mini	84.0	64.6	73.03
	Bengali prompt using Chain-of-Thought	GPT-4o-mini	92.3	78.8	85.01
	Morpheme Induction and Word Segmentation [11]	UnDivide++	86.64	80.02	83.19
NER	Bengali prompt using Zero-Shot	GPT-4o-mini	44.3	30.1	35.84
	Bengali prompt using Chain-of-Thought	GPT-4o-mini	76.0	35.8	48.67
	Bengali Named Entity Recognition using Support Vector Machine [12]	Support vector machine	89.4	94.3	91.8

However, in case of NER, traditional models like SVM still outperform GPT-4o-mini, suggesting that LLMs may require more refined prompts or additional linguistic signals for tasks involving entity classification.

Overall, the comparison underscores the growing potential of prompt engineering with LLMs as a flexible, resource-efficient alternative to conventional NLP pipelines, especially in low-resource languages.

MAP score is normally used to evaluate the performance of morphological analysis. In this paper, we have used precision, recall, and F1-score to bring uniformity in performance measurement of morphological analysis.

5.3 Error Analysis

This section presents our results on morphological analysis, named entity recognition, and part-of-speech tagging using zero-shot and chain-of-thought prompting techniques, along with an error analysis. We focus on evaluating common failure patterns, linguistic challenges, and comparative strengths/weaknesses of both approaches. Here, we divide the errors into types. For each task, look at what types of errors happen and provide examples.

5.3.1 Named Entity Recognition

In Bengali, zero-shot prompting failed to recognize even common named entities like geographic fea-

tures, such as "সমুদ্র". By encouraging the model to think about the token's meaning, even when direct language support was limited, chain-of-thought prompting improved recognition.

5.3.2 Zero-Shot Prompting Example

Let us consider an example:

Sentence: "সেখানকার জানালা দিয়ে সমুদ্র দেখা যাইতেছিল"
Zero-shot Output: No entities were recognized.
Expected Output: সমুদ্র is a Location (considered a geographic entity).

The zero-shot model failed to identify "সমুদ্র" as a location or named entity. This indicates a lack of support for Bengali or inability to handle non-Latin script without additional prompt guidance.

5.3.3 Chain-of-Thought Prompting Output

The word 'সমুদ্র' refers to a natural geographical feature. It represents a large water body and is typically treated as a location entity.

The expected output is: All tokens labeled as **"Other"**.

Although "সমুদ্র" can refer to a common noun, in this context it acts as a geographic location. The CoT prompt does not consider subtle meanings and context especially in low-resource languages like Bengali.

To improve named entity recognition (NER) in Bengali, especially in cases with implicit or context-based entities, a new prompting technique is necessary. This prompt technique struggles to capture detailed language patterns. Combining CoT with contextual templates can steer the model through organized reasoning. For effective NER, it's essential to include Bengali-specific language features like word order and pronouns.

5.3.4 Morphological Analysis

The model struggles particularly in identifying compound words and inflections. This issue stems from the model's existing dataset which does not include the specialized training required for Bengali's complex morphology. That is why it misclassified words and fails to identify root forms accurately.

Zero-shot prompting, which relies on generalized patterns, does not capture the specific nuances of the language. This situation highlights the need for tailored instruction in languages with complex morphology, like Bengali.

5.3.5 Zero-shot Prompting Output

Let us consider an example.

Sentence: "সেখানকার জানালা দিয়ে সমুদ্র দেখা যাইতেছিল"

Output:
1. সেখানকার - সেখানে - Pronoun/Determiner - Genitive case,
2. জানালা - জানালা - Noun - Singular, feminine,
3. দিয়ে - দেওয়া - Verb - Postposition (Instrumental Case),
4. সমুদ্র - সমুদ্র - Noun - Singular, masculine,
5. দেখা - দেখা - Verb - Infinitive,
6. যাইতেছিল - যাইতে - Verb - Past continuous tense.

Expected Output: দিয়ে - দেওয়া is not Postposition (Instrumental Case), it will be Verb - Present participle.

5.3.6 Chain-of-Thought Prompting Output

The results produced by Bengali Chain-of-Thought prompt gives better result than the zero-shot prompt.

Our experiments indicated that CoT prompt techniques usually improve performance by decomposing reasoning into intermediate steps.

Here, CoT did not significantly outperform the zero-shot prompting techniques for Bengali morphological analysis. Specifically, even after adding reasoning steps, some words with complex morphological structure continued to be wrongly evaluated.

5.3.7 POS-tagging Analysis

Here, we analyze the POS tagging output for a Bengali sentence "সকলেরই অনাবৃত দেহ সকলের সেই অনাবৃত বক্ষে আরশির ধুকধুকি চন্দ্রকিরণে এক একবার জ্বলিয়া উঠিতেছে" using zero-shot and Chain-of-Thought prompting techniques both.

Although most of the tokens were correctly tagged, an error was observed in the classification of the word "একবার", which was incorrectly identified as a noun.

Identified POS Tagging Error:
Word: একবার,
Expected Tag: Adverb (Temporal – meaning "once"),
Model Output: Noun,
Error Type: POS Misclassification.

While the POS tagging performance was mostly accurate, "একবার" was wrongly treated as a noun by both zero-shot and CoT prompting, which failed to categorize it as an adverb. This shows a more general problem: when the model lacks a strong linguistic basis, particularly in low-resource or morphologically rich languages like Bengali, even CoT reasoning does not always increase accuracy.

So, to improve the accuracy of POS tagging, a new prompt is required that specifically addresses the challenges of Bengali morphological nuances. The current zero-shot and Chain-of-thought prompts fail to accurately classify certain words, such as "একবার", due to their syntactic ambiguity. A more tailored approach is needed the incorporation of contextual rules or explicit instructions to handle adverbs, noun forms, and case markers.

6 Conclusion

This paper demonstrates the applicability of large language models such as GPT-4o-mini for addressing significant NLP challenges, including named entity recognition, morphological analysis, and part-of-speech tagging, in low-resource languages such as Bengali. The models showed competitive performance, frequently outperforming traditional techniques by using well-crafted prompts, especially the chain-of-thought prompting methodology. Particularly with chain-of-thought prompting, which outperformed zero-shot methods, the highest F1-scores were obtained for POS tagging, morphological analysis, and NER analysis.

The study also identifies shortcomings in LLMs' ability to handle language-specific complications, despite these encouraging findings. The difficulties encountered in NER and morphological analysis tasks highlight the need for more sophisticated, language-specific prompts that take context and morphological nuances into consideration. These nuances were not well captured by zero-shot and chain-of-thought prompting alone, especially in languages with complex inflectional systems and non-Latin scripts. More advancements are required, such as the creation of more complex and customized prompting strategies that incorporate specialized linguistic knowledge and contextual rules.

Overall, the results indicate that LLMs have a lot of potential to improve NLP tasks in low-resource languages with limited resources if they are appropriately guided by prompt engineering. To improve these models for more precise and reliable performance in linguistic analysis, more study is needed, especially for languages with complex syntax and morphology.

References

1. **Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023).** Gpt-4 technical report. arXiv preprint arXiv:2303.08774.
2. **Ashish, V. (2017).** Attention is all you need. Advances in neural information processing systems, Vol. 30, pp. I.
3. **Ashrafi, I., Mohammad, M., Mauree, A. S., Nijhum, G. M. A., Karim, R., Mohammed, N., Momen, S. (2020).** Banner: a cost-sensitive contextualized model for bangla named entity recognition. IEEE Access, Vol. 8, pp. 58206–58226.

4. **Ayman, U., Saha, C., Rahat, A. M., Khushbu, S. A. (2025).** Banglablend: A large-scale nobel dataset of bangla sentences categorized by saint and common form of bangla language. Data in Brief, Vol. 58, pp. 111240.
5. **Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020).** Language models are few-shot learners. Advances in neural information processing systems, Vol. 33, pp. 1877–1901.
6. **Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. D. O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., et al. (2021).** Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374.
7. **Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., Amodei, D. (2017).** Deep reinforcement learning from human preferences. Advances in neural information processing systems, Vol. 30.
8. **Das, A., Sarkar, S. (2019).** Morphben: A neural morphological analyzer for bengali language. International Conference on Computational Linguistics and Intelligent Text Processing, Springer, pp. 595–607.
9. **Das, P., Das, A. (2013).** Bengali noun morphological analyzer. 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, pp. 1538–1543.
10. **Dasgupta, S., Ng, V. (2006).** Unsupervised morphological parsing of bengali. Language Resources and Evaluation, Vol. 40, pp. 311–330.
11. **Dasgupta, S., Ng, V. (2006).** Unsupervised morphological parsing of Bengali. Language Resources and Evaluation, Vol. 40, No. 3–4, pp. 311–330.
12. **Ekbal, A., Bandyopadhyay, S. (2010).** Named entity recognition using support vector machine: A language independent approach. International Journal of Electrical, Computer, and Systems Engineering, Vol. 4, No. 2, pp. 155–170.
13. **Faridee, A. Z. M., Tyers, F. (2009).** Development of a morphological analyser for bengali. Proceedings of the First International Workshop on Free/Open-Source Rule-Based Machine Translation, pp. 43–50.
14. **Floridi, L., Chiriatti, M. (2020).** Gpt-3: Its nature, scope, limits, and consequences. Minds and Machines, Vol. 30, pp. 681–694.
15. **Gao, T., Fisch, A., Chen, D. (2020).** Making pre-trained language models better few-shot learners. arXiv preprint arXiv:2012.15723.
16. **Halevi, Y. (2006).** Part of speech tagging. Seminar in Natural Language Processing and Computational Linguistics (Prof. Nachum Dershowitz), School of Computer Science, Tel Aviv University, Israel.
17. **Haque, M. Z., Zaman, S., Saurav, J. R., Haque, S., Islam, M. S., Amin, M. R. (2023).** B-ner: a novel bangla named entity recognition dataset with largest entities and its baseline evaluation. IEEE Access, Vol. 11, pp. 45194–45205.
18. **Hochreiter, S., Schmidhuber, J. (1997).** Long short-term memory. Neural computation, Vol. 9, No. 8, pp. 1735–1780.
19. **Hoque, M. N., Seddiqui, M. H. (2015).** Bangla parts-of-speech tagging using bangla stemmer and rule based analyzer. 2015 18th International Conference on Computer and Information Technology (ICCIT), IEEE, pp. 440–444.
20. **Jueal Mia, M., Hassan, M., Biswas, A. A. (2022).** Effectiveness analysis of different pos tagging techniques for bangla language. Smart Systems: Innovations in Computing: Proceedings of SSIC 2021, Springer, pp. 121–134.
21. **Jurafsky, D., Martin, J. H. (2000).** Chapter 8: Word classes and part-of-speech tagging. Speech and Language Processing, Prentice Hall.

22. **Kabir, M. F., Abdullah-Al-Mamun, K., Huda, M. N. (2016).** Deep learning based parts of speech tagger for bengali. 2016 5th international conference on informatics, electronics and vision (ICIEV), IEEE, pp. 26–29.

23. **Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023).** Chatgpt for good? on opportunities and challenges of large language models for education. Learning and individual differences, Vol. 103, pp. 102274.

24. **Kumar, D., Singh, M., Shukla, S. (2012).** Fst based morphological analyzer for hindi language. arXiv preprint arXiv:1207.5409.

25. **Lample, G., Conneau, A. (2019).** Cross-lingual language model pretraining. arXiv preprint arXiv:1901.07291.

26. **Maity, S., Deroy, A., Sarkar, S. (2024).** How ready are generative pre-trained large language models for explaining bengali grammatical errors?. arXiv preprint arXiv:2406.00039.

27. **Martinez, A. R. (2012).** Part-of-speech tagging. Wiley Interdisciplinary Reviews: Computational Statistics, Vol. 4, No. 1, pp. 107–113.

28. **Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013).** Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

29. **Mohit, B. (2014).** Named entity recognition. In Natural language processing of semitic languages. Springer, pp. 221–245.

30. **Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022).** Training language models to follow instructions with human feedback. Advances in neural information processing systems, Vol. 35, pp. 27730–27744.

31. **Pakray, P., Gelbukh, A., Bandyopadhyay, S. (2025).** Natural language processing applications for low-resource languages. Natural Language Processing, Vol. 31, No. 2, pp. 183–197.

32. **Pennington, J., Socher, R., Manning, C. D. (2014).** Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543.

33. **Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019).** Language models are unsupervised multitask learners. OpenAI blog, Vol. 1, No. 8, pp. 9.

34. **Reynolds, L., McDonell, K. (2021).** Prompt programming for large language models: Beyond the few-shot paradigm. Extended abstracts of the 2021 CHI conference on human factors in computing systems, pp. 1–7.

35. **Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., Chadha, A. (2024).** A systematic survey of prompt engineering in large language models: Techniques and applications. arXiv preprint arXiv:2402.07927.

36. **Sen, O., Fuad, M., Islam, M. N., Rabbi, J., Hasan, M. K., Fime, A. A., Fuad, M. T. H., Sikder, D., Iftee, M. A. R. (2021).** Bangla natural language processing: A comprehensive review of classical machine learning and deep learning based methods. CoRR.

37. **Shahgir, H., Alam, R., Alam, M. Z. U. (2023).** Banglaconer: Towards robust bangla complex named entity recognition. arXiv preprint arXiv:2303.09306.

38. **SHANAWAZ, M. (2013).** Morphology and syntax: A comparative study between english and bangla. Unpublished master's thesis). North South University, Dhaka, Bangladesh.

39. **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017).** Attention is all you need. Advances in neural information processing systems, Vol. 30.

40. **Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022).** Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, Vol. 35, pp. 24824–24837.

41. **Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., Le, Q. V. (2019).** Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, Vol. 32.

42. **Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al. (2023).** A survey of large language models. arXiv preprint arXiv:2303.18223, Vol. 1, No. 2.

43. **Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., Irving, G. (2019).** Fine-tuning language models from human preferences. arXiv preprint arXiv:1909.08593.

Article received on 25/05/2025; accepted on 31/08/2025.
**Corresponding author is Shillpi Mishrra.*