GeoLLM: Comparative Study and Enhancement of Large Language Models for Geospatial Knowledge Using Retrieval Augmented Generation(RAG).

Serah Akojenu[1] , **Olubayo Adekanmbi**[2] and **Anthony Soronnadi**[3]
[1]Data Science Nigeria
[2]Data Science Nigeria
[3]Data Science Nigeria
{serah, olubayo, anthony}@datasciencenigeria.ai

Abstract

In recent years, large language models (LLMs) have demonstrated exceptional ability in natural language processing tasks, yet their potential to enhance geospatial intelligence remains largely untapped. Geospatial data, which is found not only in maps but also in regular text, presents unique opportunities for the application of LLMs. This research seeks to investigate if Large Language Models can serve as accurate sources of geospatial information.

To achieve this, we assessed the geospatial capability of 6 open LLMs: Mistral, GPT 3.5 and 4.o, LLaMA 2 and 3, and Gemini, using 30 curated geospatial prompts grouped into three categories: geographic coordinates, basic spatial calculations, and descriptions of places. These prompts were translated into Yoruba to ensure contextual relevance. Subsequently, we employed retrieval-augmented generation (RAG) to enhance the models' geospatial knowledge base. Our findings revealed that GPT 4.o and Mistral consistently outperformed the other models across all categories and languages, while the RAG drastically improved the LLMs' knowledge. This study demonstrates the potential of LLMs in being an accurate source of geospatial information through the use of retrieval augmented generation (RAG).

1 Introduction

The recent proliferation of Large Language Models (LLMs) has significantly advanced performance across various downstream tasks, prompting the Natural Language Processing (NLP) community to investigate the implicit knowledge embedded within their parameters[Brown *et al.*, 2020]. LLMs are increasingly regarded as comprehensive repositories of multi-domain knowledge, integrating diverse data from common sense to complex linguistic details directly into their architectures. This integration suggests that LLMs may serve as dynamic knowledge bases, incorporating a wide range of information[Safavi and Koutra, 2021], including geospatial data, which extends beyond traditional maps and charts to encompass textual descriptions of locations [Akojenu *et al.*, 2023]. Such capabilities indicate a promising avenue for LLMs to leverage their extensive knowledge in geospatial contexts[Manvi *et al.*, 2024].

Most of the world's challenges are inherently location-based, making geospatial knowledge a critical factor in making well-informed decisions across various fields and applications. The integration of geospatial intelligence enhances decision-making in areas such as disaster response, urban planning, environmental monitoring, and public health. Accessing and evaluating the geospatial reasoning capabilities of Large Language Models (LLMs) is essential to understanding their potential in processing spatial data, improving geospatial analysis, and enabling more accurate, data-driven solutions for global challenges[Mansourian and Oucheikh, 2024]. Geospatial knowledge fundamentally involves the understanding of geographic data, encompassing elements such as location, distance, and area[Bhandari *et al.*, 2023].

In this study, we aim to understand the extent of geospatial knowledge exhibited by six open-source LLMs, namely Mistral, GPT 3.5 and 4.0, LLaMA 2 and 3, and Gemini[AI, 2024][OpenAI, 2024],[AI, 2023][DeepMind, 2023], to determine their effectiveness as accurate sources of geospatial information. We also explored enhancing the enhancement of GPT-3.5's knowledge base to evaluate whether its geospatial capabilities could be improved through retrieval augmented generation (RAG). To achieve this, 30 geospatial prompts across three distinct categories were crafted: geographic coordinates, basic spatial calculations, and descriptions of places.

The English prompts were translated into Yoruba to enhance linguistic comprehensiveness, focusing on Lagos, Nigeria, as the area of interest. Lagos State was selected as the area of interest (AOI) due to its proximity to the author's location, allowing for easier verification of locations. Yoruba was chosen as the primary language since it is the most widely spoken language in the state. This proves the potential of LLMs to become sources of accurate geospatial information through RAG.

2 Literature Review

Recent studies highlight their ability to extract valuable insights from text-based geospatial data, such as climate reports and resource assessments. Koziol (2023) discusses the role of LLMs in accelerating geospatial data access through

text comprehension and natural language interfaces, making complex tasks like GIS analysis more accessible. Bhandari et al. (2023) explore LLMs' geospatial awareness by testing their ability to predict coordinates and reason about spatial relationships. Their research shows that larger models, such as LLaMA and Alpaca, perform better in geospatial tasks, confirming that LLMs encode valuable spatial knowledge and can be further enhanced through targeted fine-tuning. GeoLLM, developed by Manvi et al. (2024), demonstrates how fine-tuning LLMs with auxiliary geospatial data, such as OpenStreetMap, improves geospatial predictions like population density. This method outperforms traditional satellite-based approaches, showcasing the potential of LLMs to complement existing geospatial covariates. Large Language Models (LLMs) like GPT-3.5, GPT-4.0, LLaMA, and Mistral have shown potential in encoding geospatial knowledge.

While these studies have explored the integration of Large Language Models (LLMs) with geospatial data, they often overlook the language part, limiting accessibility and inclusivity. Most geospatial AI models and retrieval-augmented generation (RAG) techniques are primarily developed in English, excluding non-English-speaking populations from benefiting fully from such advancements. This lack of linguistic diversity reduces the effectiveness of geospatial AI in regions where indigenous languages are predominant.

3 Methodology

The study was carried out in three main phases: data curation, collection, and cleaning; comparative analysis of large language models on geospatial knowledge; and the enhancement of one of the LLM (GPT3.5).

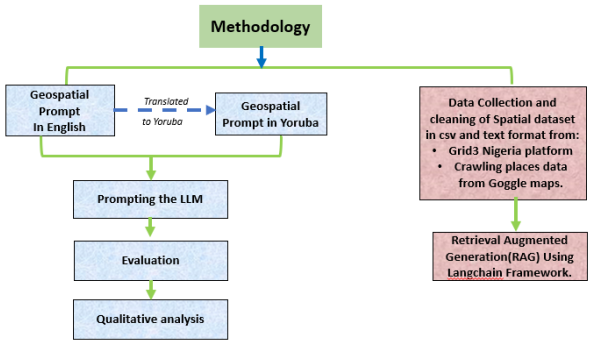

Figure 1: Methodology Flowchart

3.1 Datasets Used.

The datasets used are in two categories, the generated prompts and the Location-based dataset crawled from Google maps.

Prompts

To evaluate the geospatial capabilities of large language models (LLMs), we crafted 30 specific prompts, following the methodology outlined by [Bhandari *et al.*, 2023]. These prompts were carefully crafted to include precise location information and their answers were provided as well. After crafting the prompts, the prompts and their answers were translated into Yoruba by a professional Yoruba linguist. The prompts were organized into three primary geospatial categories as follows:

Coordinate Resolving: This category tests the ability of LLMs to convert different geographic coordinate formats to exact terrestrial locations. Examples include: Universal Transverse Mercator (UTM): For example, 448251 5411932 identifies the location of Sweet Sensation in Yaba.

Decimal Degrees (DD): For instance, 6.4854614, 3.3728524 points to the Iyana Ipaja bus stop.

Degrees and Decimal Minutes (DDM): Such as 29.128' N, 3° 22.371' E.

Basic Spatial Calculations: This category involves the LLMs' ability to execute simple spatial calculations like measuring distances or determining which locations are furthest from each other.

Description: This section assesses the LLMs' capability to provide detailed descriptions of places, emphasizing the characteristics or importance of the locations.

A sample of these prompts is seen in Table 1.

Table 1: Sample geospatial prompts in both English and Yoruba

Category	English Prompt	Yoruba Prompt
Coordinate Resolving	Where is this coordinate 6.594421, 3.345629 located in Lagos State?	Ibo ni ipò idojuk 6.594421, 3.345629 wà ní ilú Èkó?
Basic Spatial Calculation	Between iie-epo and Oshodi, which is nearer to Lekki link bridge in Lagos State?	Láàárín ilé-epo àti Oshodi, èwo ní o súnmo Lekki ní ilú Èkó?
Description	Where can I see Obafemi Awolowo statue in Lagos State?	Nibo ni mo ti lè rí ère Òbáfmi Awólw ní ilú Èkó?

Location-Based Dataset

To enhance the geospatial knowledge of the LLM, location-based datasets were utilized. This included location place data downloaded from Grid3[1] supplemented with additional location data scraped from Google Maps[2]. The datasets from Grid3 covered all of Nigeria. We filtered the data to focus on Lagos State, our location of interest, and removed irrelevant columns and non-ASCII characters. The data was primarily in .csv format. The data table for this is shown in Table 2 below;

3.2 Models

For this research, we made use of open-source models, which are easily accessible to people. Below are the large Language models considered in this research:

[1]https://data.grid3.org/

[2]https://www.google.com/maps

Table 2: Geospatial Data Categories in GRID3 and Google Maps

GRID3	Churches, energy and electricity substations, factories and industrial sites, farms, filling stations, fire stations, government buildings, health facilities, markets, police stations, schools, settlements
Google Maps	Hotels, restaurants, shops, schools, parks, churches, mosques, banks, hospitals, markets

Mistral: Mistral is a state-of-the-art AI model developed by Alphabet's Google DeepMind, offering top-tier natural language understanding and reasoning capabilities, with applications spanning multiple languages and domains, including text, audio, code, and image processing.

GPT 3.5 and 4.o: ChatGPT is a conversational AI model by OpenAI, that can answer follow-up questions, admit mistakes, challenge premises, and reject inappropriate requests. ChatGPT is similar to InstructGPT, which follows instructions and provides detailed responses. It is fine-tuned from a model in the GPT-3.5 series that was trained on an Azure AI supercomputing infrastructure. ChatGPT's research release is part of OpenAI's iterative deployment of safer AI systems, with safety mitigations based on lessons from earlier models like GPT-3 and Codex.

Llama 2 and 3: These are Language Models from Meta AI series, developed to improve upon their predecessor's capabilities in understanding and generating natural language. LLaMA 2 introduces enhanced training techniques and a larger training dataset, aimed at increasing model robustness and accuracy across diverse tasks. LLaMA 3 builds further on these advancements, incorporating state-of-the-art algorithms to refine context comprehension and response generation, thereby ensuring higher efficiency and precision in language modeling.

Gemini: Google Gemini, initially known as Bard, is a sophisticated AI chatbot tool developed by Google, leveraging natural language processing and machine learning to simulate human-like conversations, integrate seamlessly into various platforms, and perform tasks ranging from language understanding to multimodal comprehension, including text, audio, code, and video analysis, with applications across multiple domains and languages.

3.3 Comparative Analysis of the Models on Geospatial Information

The evaluation of geospatial knowledge was conducted by presenting predefined prompts in both English and Yoruba to the language models[Sikiru *et al.*, 2024]. Responses were then collected and assessed in each language, with each response being assigned a numerical weight based on its accuracy: 0 indicated an incorrect response, 0.5 denoted a partially correct or close response, and 1 signified a correct response. The accuracy of each model was subsequently calculated using the following formula:

$$\text{Accuracy} = \left(\frac{\text{Total number of correct responses (Nt)}}{\text{Total number of prompts (Np)}} \right) \times 100$$

3.4 Enhancement of Geospatial Knowledge through Retrieval Augmented Generation (RAG)

To enhance the geospatial capabilities of GPT-3.5, we employed the Langchain [3] framework by making use of an SQL agent[Li *et al.*, 2024]. This enhancement was essential to leverage our extensive datasets, which comprised numerous CSV housed within a PostgressSQL database.

4 Results and Discussion

In our work assessing language models on geographical knowledge using both English and Yoruba prompts, we discovered substantial differences in model performance between languages. To categorize the accuracy of responses, we used the following criteria:

Very Good: Highly accurate responses, such as a correct distance prediction of 34 km or a slight deviation within 34.5 km.

Good: Reasonably close estimates, such as 31 km or 32 km for an actual distance of 34 km.

Bad: Significantly incorrect responses, such as predictions of 15 km or 10 km for a correct distance of 34 km.

From the results in 3 shown below, English replies were typically more accurate, with Mistral and GPT 4.0 excelling at both languages. GPT-4.o stood out for its better performance in Yoruba, surpassing other models with an accuracy rate above 70%. However, the generally low performance in Yoruba across most models, such as Gemini, GPT-3.5, and LLaMA 2 and 3, with accuracy below 50%, highlights significant shortcomings in existing language model training methods. This emphasizes the need for more inclusive and broad training datasets that better reflect varied languages and dialects, resulting in more dependable multilingual support for applications that require precise geographic comprehension.

For the enhancement part, it was observed that with sufficient data, the knowledge and information can be drastically improved.

5 Conclusion

This study assessed the geospatial capabilities of six open-source large language models—Mistral, GPT-3.5, GPT-4.0, LLaMA2, LLaMA3, and Gemini—using curated prompts in English and Yoruba. GPT-4.0 and Mistral performed best across all categories, while models like LLaMA2 and Gemini struggled, especially with Yoruba prompts. Retrieval-augmented generation (RAG) significantly improved the models' geospatial knowledge, showing that LLMs can be effective sources of geospatial information when enhanced with external data. The findings highlight the need for more inclusive multilingual training and better geospatial datasets to improve LLM performance in diverse languages. Future work should focus on refining these models to ensure more reliable and accurate geospatial applications.

[3] https://python.langchain.com/v0.1/docs/use$_c ases/sql/quickstart/$

Table 3: Performance Evaluation

	Mistral		GPT3.5		GPT4.0		Llama2		Gemini		Llama 3	
	Eng	Yor	Eng	Yor	Eng	Yor	Eng	Yor	Eng	Yor	Eng	Yor
Very Good	63%	63%	27%	20%	73%	73%	20%	7%	43%	43%	47%	47%
Good	37%	37%	50%	37%	27%	27%	50%	37%	13%	13%	27%	27%
Bad	0%	0%	23%	43%	0%	0%	40%	90%	43%	43%	27%	27%

References

[AI, 2023] Meta AI. Llama: Open and efficient foundation language models. https://www.llama.com/, 2023.

[AI, 2024] Mistral AI. Mistral ai model. https://mistral.ai, 2024.

[Akojenu *et al.*, 2023] Serah Sessi Akojenu, Olubayo Adekanmbi, and Anthony Soronnadi. Geo-visualization of hotspots of citizens dissatisfaction on social services using media print: A case study of fuel and cash scarcity in nigeria. In *Submitted to Deep Learning Indaba 2023*, 2023. Under review.

[Bhandari *et al.*, 2023] Prabin Bhandari, Antonios Anastasopoulos, and Dieter Pfoser. Are large language models geospatially knowledgeable?, 2023.

[Brown *et al.*, 2020] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[DeepMind, 2023] Google DeepMind. Gemini ai model. https://deepmind.google/technologies/gemini/pro/, 2023.

[Li *et al.*, 2024] Jiarui Li, Ye Yuan, and Zehua Zhang. Enhancing llm factual accuracy with rag to counter hallucinations: A case study on domain-specific queries in private knowledge-bases, 2024.

[Mansourian and Oucheikh, 2024] Ali Mansourian and Rachid Oucheikh. Chatgeoai: Enabling geospatial analysis for public through natural language, with large language models. *ISPRS International Journal of Geo-Information*, 13(10):348, 2024.

[Manvi *et al.*, 2024] Rohin Manvi, Samar Khanna, Gengchen Mai, Marshall Burke, David Lobell, and Stefano Ermon. Geollm: Extracting geospatial knowledge from large language models. *arXiv preprint arXiv:2310.06213v2*, 2024.

[OpenAI, 2024] OpenAI. Gpt-4 technical report. https://openai.com/research/gpt-4, 2024.

[Safavi and Koutra, 2021] Tara Safavi and Danai Koutra. Relational world knowledge representation in contextual language models: A review. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1053–1067. Association for Computational Linguistics, 2021.

[Sikiru *et al.*, 2024] Rashidat Sikiru, Olubayo Adekanmbi, and Anthony Soronnadi. Comparative study of llms for personal financial decision in low resource language. In *AfricaNLP Workshop at ICLR 2024*, 2024.