Fostering Digital Inclusion for Low-Resource Nigerian Languages: A Case Study of Igbo and Nigerian Pidgin

Ebelechukwu Nwafor and **Minh Phuc Nguyen**
Department of Computing Sciences
Villanova University, Villanova, PA. 19085
ebelechukwu.nwafor@villanova.edu

Abstract

Current state-of-the-art large language models (LLMs) like GPT-4 perform exceptionally well in language translation tasks for high-resource languages, such as English, but often lack high accuracy results for low-resource African languages such as Igbo and Nigerian Pidgin, two native languages in Nigeria. This study addresses the need for Artificial Intelligence (AI) linguistic diversity by creating benchmark datasets for Igbo-English and Nigerian Pidgin-English language translation tasks. The dataset developed is curated from reputable online sources and meticulously annotated by crowd-sourced native-speaking human annotators. Using the datasets, we evaluate the translation abilities of GPT-based models alongside other state-of-the-art translation models specifically designed for low-resource languages. Our results demonstrate that current state-of-the-art models outperform GPT-based models in translation tasks. In addition, these datasets can significantly enhance LLM performance in these translation tasks, marking a step toward reducing linguistic bias and promoting more inclusive AI models.

1 Introduction

Machine translation (MT) is a fundamental subfield of Natural Language Processing (NLP) that involves translating text or speech between different languages using computational systems. MT has been around for a while, and its two classic methods are traditional rule-based systems and statistical machine translation (Zhang and Zong, 2020). Recently, neural machine translation (NMT) has revolutionized language translation as it uses deep learning techniques to model the translation process (Tan et al., 2020). This approach demonstrates promising results in bilingual and multilingual machine translations (Firat et al., 2016; Vaswani et al., 2017). The term Large Language Models (LLMs) is used to describe models that are built on sophisticated deep learning architectures, typically transformers, and are characterized by billions of parameters and extensive training data. These models can achieve high accuracy in various tasks – including machine translation - which is a great leap in developments in the field of natural language processing (Sarumi and Heider, 2024).

While machine translation continues to advance, the vast majority of improvements made in the last decade have been for high-resource languages, or languages that have large quantities of training data available digitally. High-resource languages such as English, German and Spanish are available in abundant resources such as books, movies, and digital texts (Pacheco Martínez et al., 2024). In contrast, low-resource languages (LRLs) are those that are less studied, resource-scarce, or less digitized, among other denominations (Magueresse et al., 2020). As of today, most NLP research focuses on a select few languages out of the vast amounts of languages spoken around the world, leaving majority of languages—particularly low-resource African languages largely understudied.

One of the major challenges to machine translation is the ability to obtain high-quality monolingual and parallel-sentence pair datasets. This ensures that the results are generated with high precision and accuracy. Unfortunately, this has been a major challenge to African language translation. This problem highlights the need to increase the linguistic diversity of African languages in the NLP fields. To address this issue, we introduce high-quality benchmark datasets in low-resource Nigerian languages - Igbo and Nigerian Pidgin - developed through a thorough process of collection, pre-processing, and annotation. In MT tasks, the parallel corpus, consists of sentences in both the source and target languages (Fan et al., 2020a). In our case, the source languages are Igbo and Nigerian Pidgin, with English as the target language. This paper presents two main contribu-

Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025), pages 44–53
May 3, 2025

tions: first, it establishes benchmark datasets for Igbo-English and Nigerian Pidgin-English machine translation tasks. We curated a combined total of over 2 million monolingual sentences for both Igbo and Nigerian Pidgin from online sources. Using this dataset, we generated parallel datasets through a combination of crowd-sourced human annotations (Amazon Mechanical Turk) and outputs from large language models (GPT-3.5 and 4). Second, it leverages the Igbo-English benchmark dataset – along with other open-source datasets in similar language pairs – to provide a comparative analysis of the translation ability of current Large Language Models (LLMs) to select state of the art machine translation models that have been shown to perform well for low-resource languages. The results demonstrate that the fine-tuned models outperform both their original version and the GPT-4 models in the Igbo-English translation task, highlighting the impact of the datasets and the effectiveness of fine-tuning on LLM performance.

2 Related Work

Recent research efforts have focused on addressing the scarcity of resources for low-resource languages by creating specialized datasets and applying them to fine-tune LLMs for translation tasks.

2.1 Datasets for Low-Resource Languages

Different approaches have been developed for the creation and augmentation of datasets. Large-scale data mining strategies and backtranslation techniques are used to construct an extensive dataset of 7.5 billion sentences for 100 languages (Fan et al., 2020a). More complex methods involve aligning language standards, hiring professional translators, conducting quality checks, and engaging independent reviewers for assessment (Costa-Jussà et al., 2022). Less costly and more common approaches include collecting data from multilingual online resources (Agić and Vulić, 2019) and employing native Igbo speakers and linguists, with inter-translator agreement and manual checks ensuring accuracy during sentence annotation (Ezeani et al., 2020). Following similar methodologies, this research contributes to the development of high-quality benchmark datasets for Igbo-English and Nigerian Pidgin-English machine translation tasks, mitigating a significant gap in the field.

2.2 Large Language Models and Translation Tasks

To improve the performance of LLMs in translation tasks, several methods have been proposed. Adding one or more language-specific layers to the final encoder of the pre-trained model has shown promise (Fan et al., 2020b). Incorporating Sparsely Gated Mixture of Expert (MoE) layers into the model has produced strong results, particularly for low-resource language translations (Costa-Jussà et al., 2022). Another approach involved tweaking the model's decoder architecture by interleaving the self-attention and cross-attention modules to better leverage the pretrained encoder before fine-tuning (Ma et al., 2021). Additionally, fine-tuning pre-trained models solely on the collected parallel corpus has been shown to significantly improve translation performance (Adelani et al., 2022).

Similarly, our research fine-tunes pre-trained models on the collected dataset without adding extra layers or modifying their architecture. The main goal is to demonstrate the effectiveness of our dataset, which has significantly boosted the Igbo translation performance of all pre-trained models, with two out of the three models surpassing GPT-4o model's performance on four benchmark datasets.

3 Data Acquisition and Preparation

In this section, we present an overview of our data preparation and acquisition process, followed by a detailed explanation of our data annotation approach. We employ the use of LLMs alongside crowdsourced native speakers as annotators to curate and annotate the dataset, ultimately creating a parallel dataset that maps the source languages (Igbo or Nigerian Pidgin) to the target language (English). [1]

3.1 Dataset

The dataset used in this work is made up of two sources. The first source consists of data we scraped from online local newspapers, broadcasters, and online dictionaries, which were then pre-processed and annotated by native speakers. The second source includes open-source benchmark datasets specifically generated for multilingual NLP tasks.

[1] The curated monolingual and parallel datasets developed from this research effort can be found at (Nguyen, 2025a) and (Nguyen, 2025b) respectively.

Data Collection. We collected data from five publicly available and credible sources: BBC Igbo [2], BBC Pidgin [3], Ted Talk [4], Voice of Nigeria (national radio broadcaster) [5], and Naijalingo (Nigerian Pidgin English dictionary) [6]. It is important to note that this dataset does not contain any personal indentifiable information. Using beautifulsoup, a python library, we scraped the aforementioned online sources and accumulated a total of over 2,000,000 monolingual sentences in both Igbo and Nigerian Pidgin.

Data Pre-processing. The dataset underwent thorough pre-processing, including line filtering, symbol removal, duplicate removal, and the elimination of redundant or little-meaning phrases, such as proper nouns, dates, and times. After pre-processing, we conducted basic data analysis on the cleaned dataset. Table 1 provides a summary of key statistics for our processed monolingual datasets.

3.2 Data Annotation approach

Crowd-sourced Human Annotation. Using Amazon Mechanical Turk [7], a crowd-sourcing platform, we curate a high-quality parallel corpus for Igbo-English and Nigerian Pidgin-English. We hired workers with the following criteria: native Nigerian speakers with a previous work acceptance rate greater than 50%. A total of 15 workers participated in the curation, 10 of them translating Igbo sentences and the remaining 5 translating Nigerian Pidgin sentences. Each worker was compensated at a rate of 4 cents per translated sentence. To avoid the use of AI or online translation tools such as Google Translate, we randomly selected and compared the original and translated sentences using Google Translate and LLMs (GPT-3.5 and 4). If we found an instance of similarity, that worker's translations would be disqualified. In addition, we scrutinized sentences that were completed in the shortest time to ensure accuracy.

LLM Annotation. In addition to human annotation, we utilize the power of LLMs for data annotation from source (Igbo and Nigerian Pidgin) to the target language (English). We generate translations for each source datasets using both GPT-3.5 and 4 models. Due to the high costs associated with utilizing LLMs, we annotated only a fraction of the original monolingual dataset. Table 2 provides a detailed summary the annotated datasets for both human and LLM annotation.

[2]https://www.bbc.com/igbo
[3]https://bbc.com/pidgin
[4]https://ted.com/
[5]https://von.gov.ng/
[6]http://www.naijalingo.com/
[7]https://www.mturk.com/

4 Experiments

In this section, we present a benchmark use case for our curated parallel datasets alongside existing datasets. Using these datasets, we evaluate the translation quality of state-of-the-art multilingual translation models optimized for low-resource languages to that of large language models. Next, we present an overview of the open-source dataset utilized in addition to our curated dataset for the comparative analysis.

4.1 Leveraging External Datasets

Flores200. This dataset is derived from English-language Wikimedia projects and includes 204 languages. Roughly one-third of the sentences come from each of the following sources: Wikinews, Wikijunior, and Wikivoyage. The data is professionally translated into over 200 languages, making Flores-200 a multilingual benchmark with many-to-many translations (Costa-Jussà et al., 2022).

MafandMT. This dataset covers 11 African languages and is sourced from news websites of local newspapers in English and French. The dataset has undergone pre-processing, translation, and quality control to ensure accuracy and reliability (Adelani et al., 2022).

JW300. This extensive dataset spans 343 languages and contains 1,335,376 articles, totaling over 109 million sentences and 1.48 billion tokens. The data is collected from a full crawl of publications on jw.org, mainly from the magazines *Awake!* and *Watchtower*. Though produced by a religious organization, the texts cover a wide range of topics and consist of translations primarily from English (Agić and Vulić, 2019).

IgboNLP. This dataset includes 5,000 English-Igbo parallel sentences collected and pre-processed from sources like Wikipedia, CommonCrawl, and local media. Translation and quality checks were performed, including manual review and inter-translator agreement (Ezeani et al., 2020).

.

Language	Source	# Sentences (Before)	# Sentences (After)	Avg Words/Sent (Before)	Avg Words/Sent (After)
Igbo	BBC Igbo	681,515	106,693	8.38	17.27
	Igbo Gov	372,100	13,265	5.18	26.63
	TedTalk Igbo	175	175	17.85	17.85
	Total Igbo	**1,053,790**	**120,133**	**7.25**	**18.3**
Nigerian Pidgin	BBC Nigerian Pidgin	1,040,775	119,225	10.17	20.7
	Naijalingo Dictionary	2,041	2,041	9.64	9.64
	Total Nigerian Pidgin	**1,042,816**	**121,266**	**10.17**	**20.5**
Total		**2,096,606**	**241,399**	**8.7**	**19.4**

Table 1: **Preprocessed Dataset Summary**. The average words per sentence reflect the complexity and density of information in sentences sourced primarily from digital newspapers.

Language	Source	# Sentences (Amazon MTurk)	Avg Words / Sentence	# Sentences (GPT-3.5)	# Sentences (GPT-4)
Igbo	BBC Igbo	12,591	21.03	104,523 (98%)	106,693 (100%)
	Igbo Gov	0	0	5,008 (37.75%)	5,008 (37.75%)
	TedTalk Igbo	0	0	175 (100%)	175 (100%)
	Total Igbo	**12,591**	**21.03**	**109,706 (91.32%)**	**111,876 (93.13%)**
Nigerian Pidgin	BBC Nigerian Pidgin	23,382	23.38	114,289 (96%)	119,225 (100%)
	Naijalingo Dictionary	0	0	2,041 (100%)	2,041 (100%)
	Total Nigerian Pidgin	**23,382**	**23.38**	**116,330 (96%)**	**121,266 (100%)**
Overall Total		**35,973**	**22.67**	**226,036 (93.64%)**	**233,142 (96.58%)**

Table 2: **Data Annotation Statistics**

5 Multilingual translation models

5.1 Model Selection

In this study, we selected NLLB, ByT5, and DeltaLM as our primary models for fine-tuning on Igbo-English translation tasks. These models were chosen based on the following factors:

Open-Source Accessibility. All three models are open-source, ensuring transparency, reproducibility, and flexibility in adapting them for low-resource language tasks. Additionally, using open-source models fosters broader community feedback and collaboration, which is particularly valuable for under-resourced languages.

Credibility of Origin. These models were developed and published by reputable institutions - Meta (NLLB), Google (ByT5), and Microsoft (DeltaLM) - which have established expertise in AI tools. Their positions as leading companies in the tech industry indicate reliability in their models.

Superior Performance in Igbo-English and similar multilingual translation task.

- **NLLB**: Demonstrated strong performance in Igbo-English translation tasks, achieving state-of-the-art results on MAFAND-MT's test set for Igbo-English translation. It outperformed previous models in both BLEU and chrF++ scores (Costa-Jussà et al., 2022).
- **DeltaLM**: Showed superior translation capabilities, achieving higher BLEU scores than mBART(Liu et al., 2020) and M2M-100 (Fan et al., 2020a) across translations from 10 diverse languages into English. Its enhanced performance makes it a compelling choice for fine-tuning on Igbo-English translation task (Ma et al., 2021).
- **ByT5**: Outperforms its predecessor, mT5, in various multilingual tasks. ByT5's strengths in handling character-level information make it especially well-suited for low-resource languages, where vocabulary-based models may face limitations (Xue et al., 2022).

5.2 Fine-tuning models

The models are fine-tuned using three open-source benchmark datasets—JW300, MAFAND-MT, and IGBONLP—along with our human-annotated dataset as detailed in the dataset section.

Hyper-parameter Tuning. The models were fine-tuned using three open-source benchmark datasets (JW300, MAFAND-MT, IGBONLP) and our human annotated dataset. Each dataset provides distinct linguistic resources for low-resource language tasks and was instrumental in model adaptation for Igbo-English translation. Specifically, the datasets include JW300 (415k sentences), MAFAND-MT (20k sentences), IGBONLP (10k sentences), and AmazonMTurk (12k sentences).

The training parameters were configured as follows: input sentences were tokenized with a maximum token length of 128. Training was conducted with a batch size of 64, while evaluation used a batch size of 8. The optimizer was set with a learning rate of 1e-4, alongside a weight decay of 0.01. The total training epochs ranged from 8 to 12, with gradient accumulation steps set to 16 to optimize memory usage. Checkpoints were saved every 150 steps, and evaluations were conducted at the same frequency to monitor model performance throughout training.

5.3 Results

Results demonstrate a significant improvement in BLEU scores for fine-tuned models on Igbo-to-English benchmark datasets. Figure 1 presents a comparison between the original models and their fine-tuned counterparts, highlighting the improvements achieved through fine-tuning. On average, NLLB's score increases by 44%, ByT5 by 17.84%, and DeltaLM by 25.64%. The original DeltaLM and ByT5 models generate poor model outputs, with BLEU scores falling below 1%. Poor performance can be attributed to insufficient Igbo training data and multilingual interference. Pre-trained on limited Igbo data, DeltaLM and ByT5 models struggle to capture the language's unique linguistic patterns, resulting in mixed-language outputs during translation. Fine-tuning with Igbo-specific data effectively addresses these challenges, leading to significant performance improvements.

Furthermore, we evaluated the BLEU scores of five different models across four different datasets. The results show the performance variations for both original and fine-tuned versions of the models. Figure 2 presents a BLEU score comparison across various benchmark datasets for state of the art models and LLM (GPT-4o). The results indicate that the fine-tuned NLLB achieves the highest performance on the Flores200 and AmazonMTurk datasets, while fine-tuned DeltaLM outperforms others on MafandMT and JW300. Overall, the

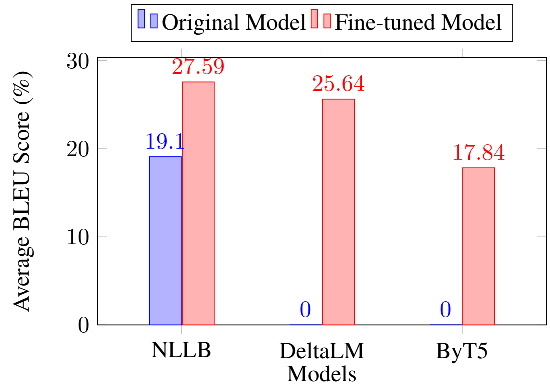


Figure 1: **Performance Comparison of Original and Finetuned Models**

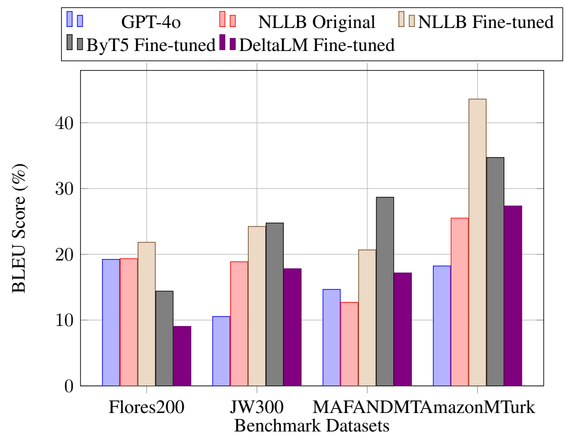


Figure 2: **BLEU Score Comparison Across Datasets for Different Models**

fine-tuned NLLB attains the highest average BLEU score, ByT5 and DeltaLM Fine-tunned models perform well in certain datasets, while GPT-4o underperforms when compared to the fine-tuned models in most of the datasets).

6 Conclusion and Future Work

In this paper, we present benchmark datasets for two low-resource Nigerian languages: Igbo and Nigerian Pidgin. We demonstrate the use case of the dataset by leveraging the Igbo benchmark datasets, combined with other open-source benchmark datasets, to significantly improve state-of-the-art translation models for low-resource languages. Our result significantly boosts the models' performance in low-resource language translation tasks, with BLEU scores exceeding those of the baseline models, including GPT-4o. This result demonstrates the effectiveness of fine-tuning for low-resource languages.

Beyond the technical contribution, our research plays an important role in the diversification of AI datasets by promoting linguistic inclusivity. The establishment of robust benchmark datasets for Igbo and Nigerian Pidgin provides a critical foundation for future research and development, helping reduce linguistic bias and expanding the scope of machine translation to underrepresented languages. This work not only underscores the potential of LLMs in low-resource settings but also opens new avenues for creating more equitable AI applications across languages.

References

David Ifeoluwa Adelani, Jesujoba Oluwadara Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Chinenye Emezue, Colin Leong, Michael Beukman, Shamsuddeen Hassan Muhammad, Guyo Dub Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ayoade Ajibade, Tunde Oluwaseyi Ajayi, Yvonne Wambui Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Koffi Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. 2022. A few thousand translations go a long way! leveraging pre-trained models for african news translation. *Preprint*, arXiv:2205.02022.

Željko Agić and Ivan Vulić. 2019. JW300: A wide-coverage parallel corpus for low-resource languages. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3204–3210, Florence, Italy. Association for Computational Linguistics.

Marta R Costa-Jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Ignatius Ezeani, Paul Rayson, Ikechukwu Onyenwe, Chinedu Uchechukwu, and Mark Hepple. 2020. Igbo-english machine translation: An evaluation benchmark. *Preprint*, arXiv:2004.00648.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. 2020a. Beyond english-centric multilingual machine translation. *Preprint*, arXiv:2010.11125.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. 2020b. Beyond english-centric multilingual machine translation. *CoRR*, abs/2010.11125.

Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016. Multi-way, multilingual neural machine translation with a shared attention mechanism. *Preprint*, arXiv:1601.01073.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Preprint*, arXiv:2001.08210.

Shuming Ma, Li Dong, Shaohan Huang, Dongdong Zhang, Alexandre Muzio, Saksham Singhal, Hany Hassan Awadalla, Xia Song, and Furu Wei. 2021. Deltalm: Encoder-decoder pre-training for language generation and translation by augmenting pretrained multilingual encoders. *Preprint*, arXiv:2106.13736.

Alexandre Magueresse, Vincent Carles, and Evan Heetderks. 2020. Low-resource languages: A review of past work and future challenges. *CoRR*, abs/2006.07264.

Tommy Nguyen. 2025a. Igbo & nigerian pidgin monolingual datasets. https://github.com/Tommy0201/WebScraped-LowResource-Translation-Dataset.

Tommy Nguyen. 2025b. Igbo & nigerian pidgin parallel datasets. https://github.com/Tommy0201/LowResource-Paper-Dataset.

Máximo Enrique Pacheco Martínez, Maya Carrillo Ruiz, and María de Lourdes Sandoval Solís. 2024. Harmony search for hyperparameters optimization of a low resource language transformer model trained with a novel parallel corpus ocelotl nahuatl – spanish. *Systems and Soft Computing*, 6:200152.

Oluwafemi A. Sarumi and Dominik Heider. 2024. Large language models and their applications in bioinformatics. *Computational and Structural Biotechnology Journal*, 23:3498–3505.

Zhixing Tan, Shuo Wang, Zonghan Yang, Gang Chen, Xuancheng Huang, Maosong Sun, and Yang Liu. 2020. Neural machine translation: A review of methods, resources, and tools. *AI Open*, 1:5–21.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Linting Xue, Aditya Barua, Noah Constant, Rami Al-Rfou, Sharan Narang, Mihir Kale, Adam Roberts, and Colin Raffel. 2022. Byt5: Towards a token-free future with pre-trained byte-to-byte models. *Preprint*, arXiv:2105.13626.

Jiajun Zhang and Chengqing Zong. 2020. Neural machine translation: Challenges, progress and future. *Preprint*, arXiv:2004.05809.

A Appendix

A.1 Sampling translations by MTurk workers, GPT 3.5, GPT 4

A.2 Amazon Mechanical Turk Worker Instructions

Workers were provided with a comprehensive set of instructions, including a task summary, detailed guidelines, and illustrative examples to ensure clarity and consistency in their annotations. Figures 3 & 4 provides a summary of the instruction provided to the annotators.

A.3 Translation Prompt Design

The following prompt was designed for data annotation using GPT-3.5 and GPT-4. It instructs the models to translate Igbo sentences into English with precision and accuracy.

- **System role:** *You are a professional translator specializing in Nigerian languages. Your task is to translate Igbo sentences into English.*

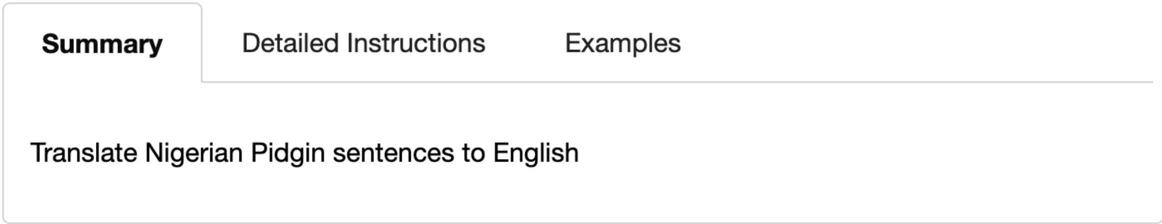
- **User role:** *Translate this sentence to English:* {sent}. *Please provide the most accurate translation possible. Only include the translated sentence in your response.*

Source Sentence (Igbo)	MTurk Worker Translation	GPT-3.5 Translation	GPT-4 Translation
Etu ahụ kwa ka Liverpool si merie ndị Manchester City.	That's how Liverpool beat Manchester City.	The match ended in a draw between Liverpool and Manchester City.	That's how Liverpool defeated Manchester City.
Ọsọ ka m ji gbapụta na-eti mkpu na-akpọ ndị mmadụ ka ha bịa nyere m aka.	I ran out shouting and calling on people to come and help me.	It is rare for me to receive a gift from people willingly.	I ran out screaming, calling for people to come and help me.
Anyị anaghị esekpu were ekele Igwe, ọ bụghị omenala Igbo.	We don't greet Igwe, it's not Igbo culture.	We do not take for granted the blessings of the Lord, it is not an Igbo tradition.	We do not kneel down to greet the King, it is not Igbo tradition.
Nkọwa foto, Uche Nwosu bụ onyeisi ndị ọrụ na Imo steeti.	Photo caption, Uche Nwosu is the chief of staff in Imo state.	Uche Nwosu is the head of workers in Imo state.	In the photo, Uche Nwosu is the leader of workers in Imo state.
Anyi ga-ezute n'ụlọ oriri na ọṅụṅụ.	Most of them are optimistic that they will win an award for their efforts in various movies last year.	All of them are determined to succeed and they will strive for excellence because they have experienced various challenges in the past year.	Many of them are hopeful that they will win awards for their efforts in various film productions last year.

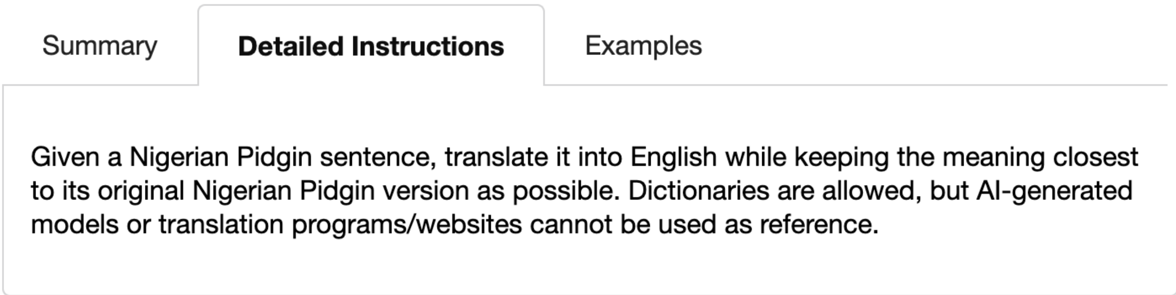
Table 3: **A sample of Igbo to English translations by MTurk workers, GPT-3.5, and GPT-4.**

Source Sentence (Nigerian Pidgin)	MTurk Worker Translation	GPT-3.5 Translation	GPT-4 Translation
Di execution of di for di capital high-security Tihar prison na di first for India since 2015.	The execution of the for the capital high-security Tihar prison is the first in India since 2015.	The execution of the high-security Tihar prison in the capital is the first in India since 2015.	The execution of the death sentence for the high-security Tihar prison in Delhi is the first in India since 2015.
Police no know wen di last Dapchi girl go free.	Police have no idea on when the last Dapchi girl will be set free.	The police do not know when the last Dapchi girl will be released.	The police do not know when the last Dapchi girl will be freed.
Why dis billionaire wey get fashion retail Patagonia dey dash im company to charity.	Why this billionaire who owns fashion retail Patagonia is giving out his company to charity.	Why is this billionaire who owns the fashion retail Patagonia giving away his company to charity?	Why is this billionaire who owns the fashion retail Patagonia giving his company to charity?
Ghana students vex over 40% increase in public universities school fees.	Ghana students are furious over the 40% increase in public universities school fees.	Ghanaian students are angry about a 40% increase in public universities' school fees.	Ghanaian students are angry over a 40% increase in public university tuition fees.
Strike dey come if minimum wage no come - Trade unionist.	We will go on strike if they don't pay minimum wage - Trade unionist.	A strike is imminent if the minimum wage is not implemented - Trade unionist.	A strike is coming if the minimum wage does not increase - Trade unionist.

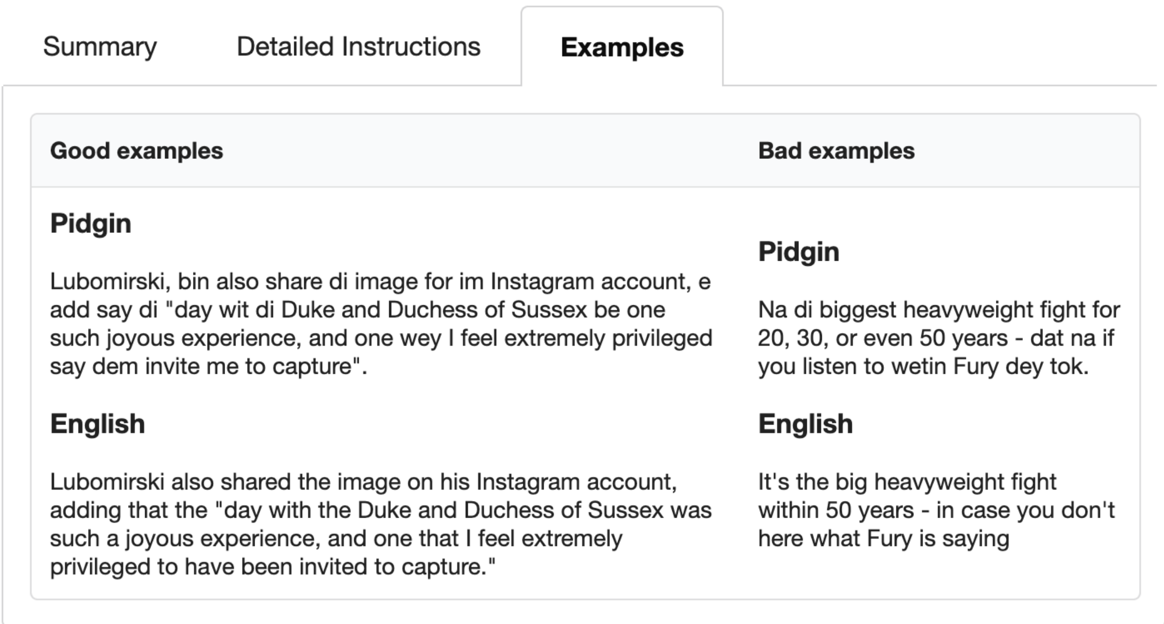
Table 4: **A sample of Nigerian Pidgin to English translations by MTurk workers, GPT-3.5, and GPT-4.**



(a) Task summary provided to translators.



(b) Detailed instructions of the annotation process.



(c) Examples of translated sentences.

Figure 3: Overview of the annotation task, including task summary, detailed instructions, and examples of Pidgin to English translated sentences.

Summary | Detailed Instructions | Examples

Translate Igbo sentences to English. If your work got accepted and you got paid, you are qualified for the much larger but similar HIIT tasks of translating 2,500 sentences

(a) Task summary provided to translators.

Summary | **Detailed Instructions** | Examples

Given an Igbo sentence, translate it into English while keeping the meaning closest to its original Igbo version as possible. Dictionaries are allowed, but AI-generated models or translation programs/websites cannot be used as reference.

(b) Detailed instructions of the annotation process.

Summary | Detailed Instructions | **Examples**

Good examples	Bad examples
Igbo Mịnịsta na-ahụ maka njem na-ahụ maka njem, Muazu Sambo ekwughachila atụmatụ gọọmentị Naịjịrịa iwepụta nde mmadụ iri abụọ na otu ọrụ oge niile na iwepụta nde mmadụ 35 n'ogbenye n'afọ 2025….. **English** The Minister of Transportation, Muazu Sambo, has reiterated the Nigerian government's plans to allocate 21 million full-time jobs and to deliver 35 million people out of poverty by 2025.....	**Igbo** Nwa ada a dị afọ 27 chọrọ idebanye aha ya nakwa aha Naịjịrịa n'akwụkwọ 'Guiness Book of Records' dịka onye kacha sie nri ogologo oge n'ụwa niile.. **English** The child, at the age of 27, broke the record by registering his name in the Guinness Book of Records as the youngest self-made millionaire in the world.

(c) Examples of translated sentences.

Figure 4: Overview of the annotation task, including task summary, detailed instructions, and examples of Igbo to English translated sentences.