Lost in Translation: Translating Generation Z Internet Slang using Machine Learning

A Special Problem Proposal
Presented to
the Faculty of the Division of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Visayas
Miag-ao, Iloilo

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science by

FLAUTA, Neil Bryan
GIMENO, Ashley Joy
GIMENO, Carl Jorenz

Francis DIMZON
Adviser

April 30, 2025

Abstract

Internet slang is an informal variation of language that is prominent to the younger generation. The usage of this language brought a generational divide between them and the older generations. This study aimed to develop a translation tool leveraging Large Language Models (LLMs) to bridge this issue. A dataset of Generation Z slang sentences and their formal equivalents was used to fine-tuned Zephyr-7B-Beta model. The performance of the fine-tuned model was evaluated against the base model using automatic metrics (BLEU and ROUGE-L) and manual evaluations through online surveys involving Gen Z students. Results showed that the fine-tuned model only slightly outperformed the base model in terms of automatic metrics, and it was generally preferred by human evaluators. These results indicate the fine-tuned model's effectiveness in producing more contextually appropriate and user-aligned formal translations.

Keywords: Internet Slang, Generation Z, Generational Divide, LoRA, LLM

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Language is how humans communicate and express themselves (Crystal & Robins, 2024). It evolves, adapting to the changing needs of users (Jeresano & Carretero, 2022). New words are borrowed or invented (Mantiri, 2010), and most linguistic changes are initiated by young adults and adolescents (Thump, 2016 as cited in (Jeresano & Carretero, 2022)). This demographic tends to focus on belonging to self-organized groups of peers and friends, forming what can be described as the "we" generation. Through their interactions, language changes differently, making them remarkably distinct from previous generations.

Slang is a great example of the dynamic nature of language. Slang is an informal language used by people in the same social group (Fernández-Toro, 2016). It serves multiple social purposes: identifying group members, communicating informally, and opposing established authority (McArthur, 2003). Slang is highly contextual and pervasive, even in non-standard English. Its figurative nature and how it twists the definitions of the words used make it difficult for outsiders to understand.

In recent years, the Internet has become a significant medium for the evolution and spread of language, giving rise to 'Internet slang' (J. Liu, Zhang, & Li, 2023). Internet slang is a collection of everyday language forms used by various online groups (Barseghyan, 2014). Ujang et al. (2018, as cited in (binti Sabri, bin Hamdan, Nadarajan, & Shing, 2020)) state that internet slang is not easily understood by people outside the social group or people who are not fluent in the language where the slang is used. This phenomenon is particularly prominent among the younger generation (Maulidiya, Wijaya, Mauren, Adha, & Pandin, 2021), where they use it to communicate and interact with friends.

Generation Z, individuals born between 1996 and 2009, are regarded as "digital natives" because technology is an integral part of their upbringing (Dua et al., 2024). Even the language of this generation is greatly affected by technology, where newly coined terms and phrases, called Gen Z slang, are tied to the media culture they've grown up with (Jeresano & Carretero, 2022). However, this evolution of language often creates communication barriers with older generations (Venter, 2017 as cited in (Ghazali & Abdullah, 2021)). Furthermore, studies show that even within Generation Z, people with limited exposure to social media may struggle to understand the prevalent slang (Vacalares, Salas, Babac, Cagalawan, & Calimpong, 2023).

These gaps highlight the need for a tool that can bridge the generational divide, making it easier for individuals to understand the language of Generation Z. Multiple studies have tried translating slang into a formal language using machine learning. Khazeni et al. achieved a 81.91% accuracy in translating Persian slang to formal Persian language using deep learning. Another study by Nocon et al. created a translator to translate Filipino colloquialisms into the Filipino language using Tensorflow's sequence-to-sequence model and Moses' phrase-based statistical machine translation. Furthermore, Ibrahim and Sharief developed a slang translator using models from Hugging Face.

Building on these studies, this study proposes to create a translation tool specifically to translate Gen Z slang. The tool will utilize Low Rank Adaptation (LoRA) to a selected Large Language Model (LLM). The results will be evaluated using the Recall-Oriented Understudy for Gisting Evaluation (ROUGE).

By fostering mutual understanding, this tool aims to promote more effective and harmonious interactions across age groups, ultimately enhancing relationships and reducing miscommunication.

The main contributions of this study are as follows:

- Enhance linguistic understanding between generations by using fine-tuning a LLM to translate Gen Z slang to formal language, leveraging the strengths of advanced NLP techniques
- Bridge communication gaps between generations using the proposed model to foster better relationships
- Create a scalable framework that can be adapted to translate slang in other languages

1.2 Problem Statement

Internet slang fosters informal, relatable communication within the younger generation (Ghazali & Abdullah, 2021), especially Generation Z, but it presents challenges in understanding for people outside this demographic. The gap in comprehension with older generations widens as internet slang evolves, often leading to miscommunication affecting social relationships that contribute to the generational divide (Vacalares et al., 2023). A more specific translation tool developed using language models can be used to bridge this divide.

By leveraging the ability of LLM to generate a more nuanced and properly constructed answer, a better tool can be made to translate the slang into proper sentences. It has already been proven by the likes of GPT being modified and tailored for use in several automated chatbots to provide customer service.

1.3 Research Objectives

1.3.1 General Objectives

This study aims to fine-tune the zephyr-7b LLM for use in the translation of Generation Z internet slang used by Filipinos in social media.

1.3.2 Specific Objectives

- To create a dataset of sentences containing Generation Z slang used in differing contexts and its formal translation
- To create a LoRA implementation for fine-tuning an existing model
- To fine-tune an existing LLM to translate sentences containing Generation Z slang into formal sentences
- To evaluate the performance of the trained model and compare it to the baseline model using several performance metrics

1.4 Scope and Limitations of the Research

This study focused on the use of internet slang by Filipino Generation Z, with an emphasis on the English language, as it is widely used on different digital platforms, such as social networks.

1.5 Significance of the Research

The study contributed to understanding the evolving linguistic landscape shaped by Internet slang, especially as used by Generation Z. The insights gained from this study aid educators, parents, and communication professionals in bridging inter-generational communication gaps and fostering better understanding across age groups.

Chapter 2

Review of Related Literature

2.1 Communication Gap between Generations

Language is dynamic in nature and thus, constantly evolving over time. One example of this behavior is the development of internet slang. Internet slang is a result of language variation and is often regarded as informal (S. Liu, Gui, Zuo, & Dai, 2019). In the study, *The Use of Online Slang for Independent Learning in English Vocabulary* (Ambarsari, Amrullah, & Nawawi, 2020), students used internet slang to express their feelings and emotions, and to align their communication style with their peers.

However, this development has its challenges. It is suggested that younger generation should use slang to communicate with each other instead of older generations because it might cause confusion between them (Jeresano & Carretero, 2022).

This miscommunication is prominent between generations with differences in linguistic familiarity as Suslak (Suslak, 2009) argues that age influences language use, noting that language evolves across generations. Supporting this, a study by Teng and Joo (Teng & Joo, 2023) found that the older a person is, the less likely they are to understand internet language.

Studies have shown that using internet slang improves relationships between those who use it. However, using internet slang for inter-generational communication can be a hindrance to proper and effective communication (Gonzaga, 2025).

2.2 Generative AI

Generative AI encompasses machine learning models that create new content, such as text, images, and audio, based on patterns learned from extensive data (Euchner, 2023). These models, including LLMs like those used in ChatGPT and Bing AI, use neural networks to predict the next word or phrase in a sequence, enabling them to generate human-like text (Brynjolfsson, Li, & Raymond, 2023). The ability of generative AI to understand and produce diverse content, ranging from creative writing code, makes it potentially useful for various applications, such as language translation (Fui-Hoon Nah, Zheng, Cai, Siau, & Chen, 2023).

2.3 Existing Studies

Vergho et al. (Vergho, Godbout, Rabbany, & Pelrine, 2024) used multiple open source LLMs and compared them with the latest ersion of GPT-3.5 and 4.0 models at that time. They determined zephyr-7b-beta is a viable open-source alternative to these models and is comparable with the latest GPT-4.0 model.

Khazeni et al. (Heydari, Albadvi, & Khazeni, 2024) used deep learning to create a model for translating Persian slang text into formal ones. The researchers explored the challenges of translating Persian slang into English within the context of film subtitling, specifically focusing on the performance of three neural machine translation (NMT) systems, namely Google Translate, Targoman, and Farazin. The primary interest of the paper lies in the understanding of how these NMT systems handle the complexities of slang translation. It was revealed that the NMT systems often struggle to capture the nuances of slang, leading to unnatural and inaccurate translations. Targoman performed best in naturalness, but it fell short of human translation quality. This implies the need for specialized algorithms or training data suitable for slang, and potentially human post-editing, to achieve accurate and culturally appropriate translations in this domain.

The study by Nocon et al. (Nocon, Kho, & Arroyo, 2018) explores translating Filipino colloquialisms, such as Conyo and Datkilab, into standardized Filipino, addressing comprehension barriers for non-familiar speakers. Two machine translation (MT) approaches were evaluated: Tensorflow's Sequence-to-Sequence model using Recurrent Neural Networks (RNNs) and Moses' Phrase-based Statistical MT. Moses outperformed Tensorflow on test data due to its handling of phrase combinations and unfamiliar words, while Tensorflow excelled on training data, indicating potential with refinement and more training data. The research underscores the need for robust datasets and highlights the strengths of phrase-based statistical MT in tackling slang translation challenges.

Ibrahim and Mustafa (Ibrahim & Sharief, 2023) developed a system to translate slang into formal language, addressing challenges posed by slang's informality and variability. Using updated datasets of slang words, formal equivalents, and contextual sentences, they fine-tuned pre-trained models from Hugging Face's Transformer library. While the T5-base model showed promise during training, it performed poorly in testing. In contrast, the "facebook/bart-base" model excelled, demonstrating high accuracy and low loss values. The study highlights the importance of fine-tuning and updating datasets for effective slang translation and emphasizes the potential of transformer models like "facebook/bart-base" in bridging informal and formal language gaps.

2.4 LoRA for Fine Tuning

Low Rank Adaptation, or LoRA, is an efficient Parameter Efficient Fine Tuning (PEFT) method proposed by Hu et al (Hu et al., 2021). This can significantly decrease the required storage for training while producing comparable results and in some cases even outperforming other adaptation methods. In addition, it has minimal chance of catastrophic forgetting as the original weights are not being tampered with, unlike other fine-tuning methods. These factors make it a suitable option for slang translation as a quick yet accurate solution. In a study conducted by Zhao et al. (Zhao et al., 2024), they determined that some LLMs using LoRA for fine tuning can outperform GPT-4, one of the most advanced LLM models currently. A study by Nguyen et al. (Nguyen, Wilson, & Dalins, 2023) used LoRA in fine tuning a pre-trained Llama 2 7B model for text classification of a dataset that contains slang. They were able to create a more accurate model compared to models by existing studies at that time.

2.5 Chapter Summary

This chapter shows how generational differences create communication gaps, especially due to internet slang. Younger people tend to use slang to express emotions and connect with friends, but this can confuse older generations who aren't as familiar with these terms. Research shows that as language changes over time, older people are generally less likely to understand the newest internet language. To bridge this gap, some recent studies have utilized machine learning to translate slang into more standard language. For instance, Khazeni et al. (Heydari et al.,

2024) used deep learning to translate Persian slang, while Nocon et al. (Nocon et al., 2018) created a Filipino slang translator using statistical models. Moreover, Ibrahim and Mustafa (Ibrahim & Sharief, 2023) fine-tuned pre-trained models to learn slang meanings. One promising technique for this is Low Rank Adaptation (LoRA), which is a fine-tuning method that keeps the original model stable while using less storage. Studies by Zhao et al. (Zhao et al., 2024) and Nguyen et al. (Nguyen et al., 2023) show that LoRA models are not only efficient but can even outperform advanced models like GPT-4 when it comes to slang translation and text classification.

Table 2.1: Summary of Existing Studies

Author	Focus	Gaps	Problem Solved
Nocon et al.	Developing machine translators for Filipino colloquialisms using sequence-to-sequence models and statistical machine translation (Moses).	Tensorflow models had issues with unknown tokens and repetitions, and limited ability to generalize to unseen data.	Demonstrated the feasibility of machine translation for Filipino colloquialisms, with Moses as a viable solution.
Ibrahim et.al	Developing an intelligent system to transform English slang words into formal words.	The study noted that more powerful processors could improve the training and testing, and that previous datasets were outdated and needed updating.	Demonstrated an effective model for translating English slang to formal English and highlighted the importance of fine-tuning pre-trained models.
Khazeni et al.	Persian slang text conversion to formal and deep learning of Persian short texts on social media	The BERT models used did not align well with the informal data used in the sentiment analysis.	Created a tool to convert Persian slang to formal text and improved sentiment analysis of short texts using deep learning.

Chapter 3

Research Methodology

This chapter lists and discusses the specific steps and activities that will be performed to accomplish the project. The discussion covers the activities from pre-proposal to Final SP Writing.

3.1 Research Activities

3.1.1 Data Gathering

A dataset of sentences containing Generation Z slang and its formal translation was used in this study. This dataset was created using several source: data obtained from social media posts and manually translated by the researchers, existing datasets from HuggingFace, and machine generated and translated sentences using GPT-4o from OpenAI.

The data obtained from social media posts were from verified users of X whose ages are within the Generation Z, so that the dataset is accurate. The data was manually translated by the researchers to ensure that the translation is accurate and reflective of the target demographic. Data obtained from existing datasets and GPT-4o was checked manually to check if whether the sentence is one used by Generation Z. These processes ensured that the dataset is of high quality and representative of what and how Generation Z slang is used.

3.1.2 Data Preprocessing

The dataset used for the fine-tuning of the model was preprocessed to ensure optimal performance of the model. Unnecessary information such as email addresses and URLs was removed. The data was then manually cleaned up to remove unnecessary characters such as emojis and fixed issues such as typos. A similar approach was done with existing and machine generated datasets to ensure consistency within the training dataset.

The dataset is then split into train and test datasets in a 90/10 ratio to maximize the data learned by the model without compromising on the model's ability to generalize to new data. The train dataset is then split again into a 90/10 ratio to ensure no overfitting while still allowing the model to adapt to the pattern of slang. The cleaned up dataset was then tokenized through the Transformers library provided by HuggingFace as the library already has tokenizers available for their pretrained models. This ensures that the data is formatted properly as required by the model to be used.

3.1.3 Model Fine-Tuning

The model used in this study was zephyr-7b-beta because it is open-source and was proven to perform better than other models of the same size. In addition, it can be trained in a GPU with 16GB of VRAM, necessary as we are using the free tier of Google Colab as the platform of choice for prototype fine-tuning of the model.

This study used the example codes provided by HuggingFace in the documentation of their various libraries and sample notebook provided in the zephyr-7b-beta repository.

The model was loaded using the Transformers library and was quantized into 4 bits through BitsandBytes library to fit the entire model in the allocated resources while having enough headroom for training. In addition, the Unsloth library was used to speed up the training time and reduce the resources used even more (Daniel Han & team, 2023). A LoRA adapter was then attached to the model to further reduce the parameters to be trained.

To evaluate the model training process and ensure that the model is not overfitting, Bilingual Evaluation Understudy (BLEU) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) are used. BLEU is used to measure the precision of the model by determining how much of the generated text appear in the reference text (Papineni, Roukos, Ward, & Zhu, 2001) while ROUGE is used to measure recall as it determines how much of the reference text is in the generated text (Lin, 2004). These metrics use n-grams, making them superior to standard recall and precision metrics as they take into account the positioning of the words. These two metrics were implemented using the Evaluate library by HuggingFace, making it easier to integrate with the rest of the model training process. These metrics was calculated at every epoch of the training process and is used for an early stopping callback to immediately stop the model training if the model seems to be overfitting.

The model was then trained using SFTTrainer from the TRL library of HuggingFace to simplify the training process. The model was trained with the following parameters: optimizer is paged 4bit AdamW, batch size of 8, learning rate of 2e-5, and maximum number of epochs of 50. These parameters were chosen based on the GPU provided in Colab, the test notebook by HuggingFace and the default parameters of SFTTrainer.

3.1.4 Model Evaluation

The model was evaluated using both automatic and manual evaluation metrics. The model was then prompted to generate a formal sentence for each sentence in the test dataset. The generated sentences were then compared to the formal translation of the sentence using BLEU and ROUGE metrics. The base zephyr-7b-beta model was also prompted to generate sentences for the BLEU and ROUGE metric and the pairwise comparison for human evaluation. Identical answers between the finetuned and the base model were removed to in the test set to ensure that the model is evaluated properly. A total of 144 sentences were used to evaluate the model.

A survey was conducted to compare the finetuned model to the base model to determine if the finetuning was effective. The survey was conducted online using Google Forms asked the participants to pick which of the following sentences is the more accurate translation of the given sentence based on accuracy, naturalness, and context. The order in which sentences from the two models were shown was randomly selected to avoid bias. To improve the response rate of the survey, the survey was split into multiple sets, answered by the same groups of people, allowing them to answer any or all of the survey forms.

3.2 Calendar of Activities

Table 3.1 shows a Gantt chart of the activities. Each bullet represents approximately one week's worth of activity.

Table 3.1: Timetable of Activities

Activities (2024-2025)	Dec	Jan	Feb	Mar	Apr	May	Jun
Creation of the dataset	•						
Identification of potential LLM to be used	•						
Lookup on available GPU on demand services	•						
Study on LoRA implementation for LLM	•						
Preprocessing of data	•••						
Prototype implementation of LoRA	•	••••					
Implementation of LoRA on selected model			••				
Implementation of LLM Evaluation Metrics			••				
Model Evaluation and Analysis of Results				••••			
Documentation	••	••••	••••	••••	••••		

Chapter 4

Results & Discussions

4.1 Dataset

We built a dataset containing a total of 1155 Gen Z internet slang sentences and their corresponding formal translations. The created dataset was then combined with another dataset from Hugging Face that contains 548 Gen Z internet slang and their corresponding formal translation.

4.2 Model Evaluation

4.2.1 Model Training

The model was trained for 7 epochs before the early stopping callback was triggered because the evaluation metrics has not improved by at least 0.01 for 3 consecutive epochs. This prevented the overfitting seen in the following figure.

Here, we can see that the while the training loss is decreasing, the validation loss is increasing and other metrics are not improving. This indicates that the model is overfitting to the training data and may not generalize well to new data. The model training was stopped in just 7 epochs and the best model amongst the epochs, the one with the lowest validation loss and highest metrics, was chosen as the final model.

Figure 4.1: Training Loss

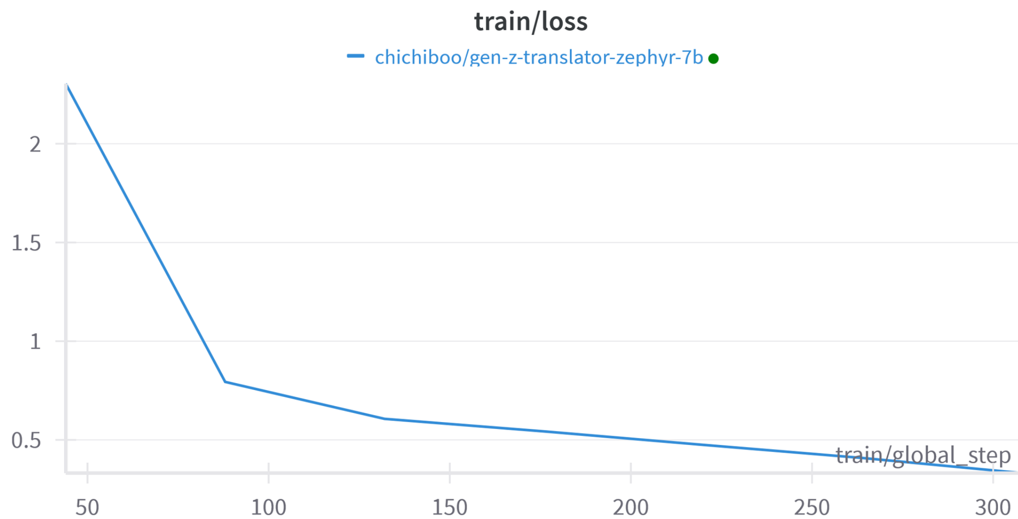


Figure 4.2: Validation Loss

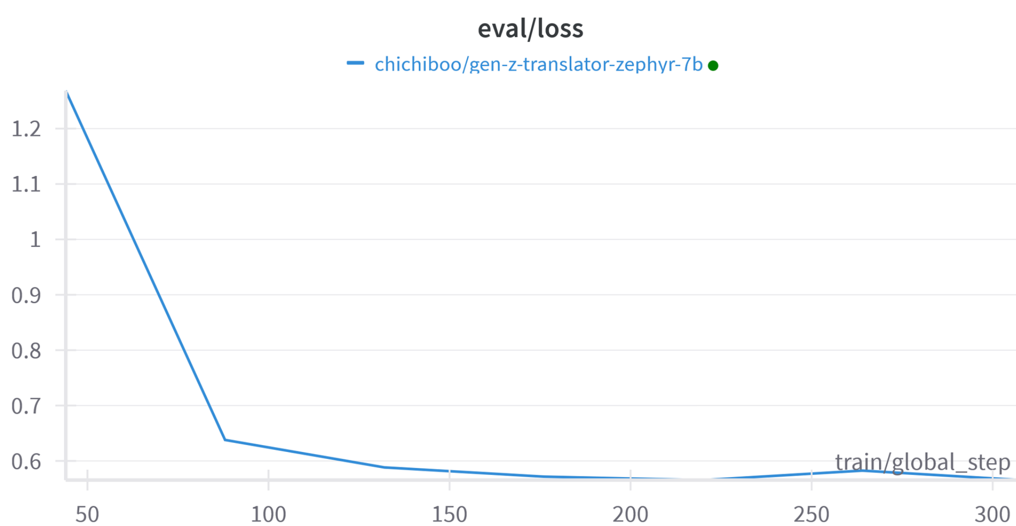


Figure 4.3: Evaluated using BLEU metric

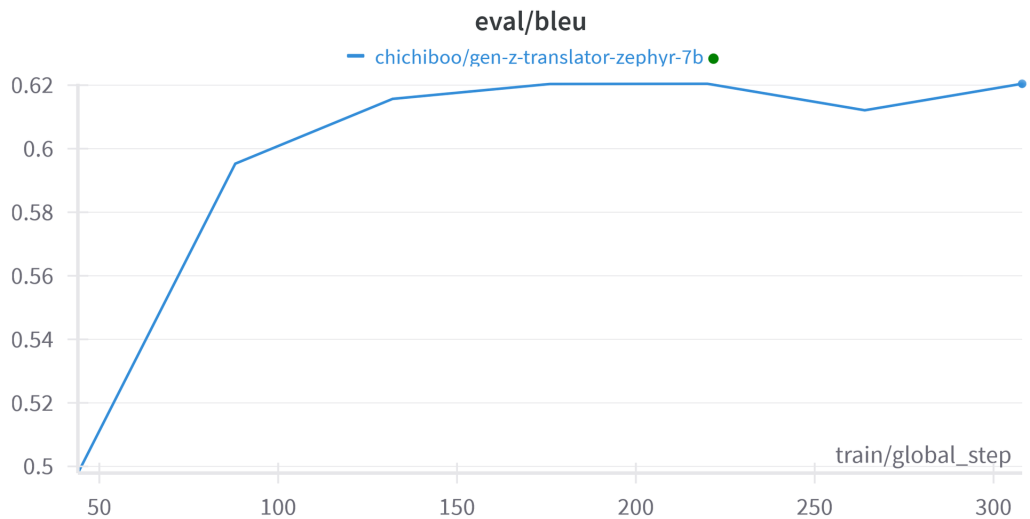
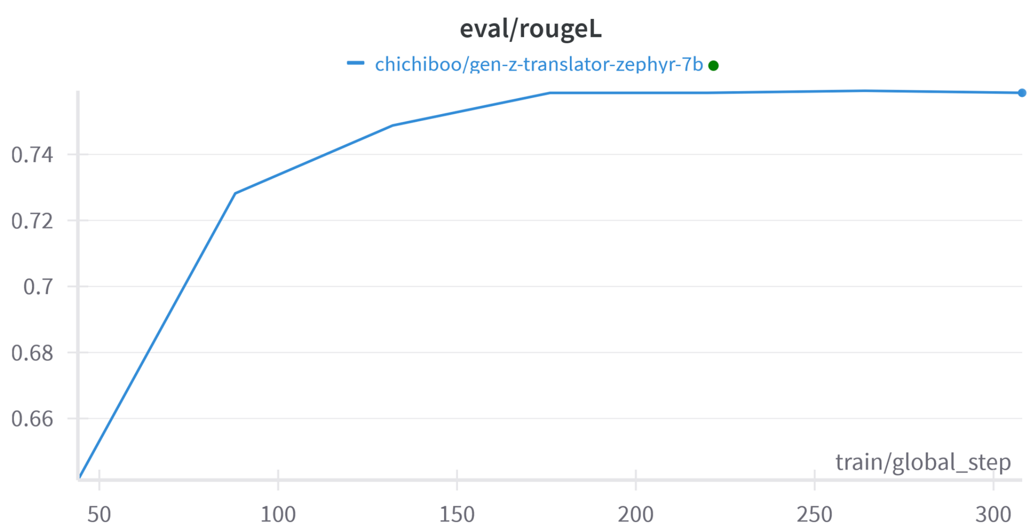


Figure 4.4: Evaluated using ROUGE-L metric



4.2.2 Text Generation

A total of 197 sentences were translated using both the base zephyr-7b-beta model and the finetuned model. These served as the dataset used to evaluate the performance of the model and comparing it with the other base model.

4.2.3 Automatic Evaluation Metrics

The dataset was automatically evaluated using BLEU and ROUGE metrics, specifically the ROUGE-L metric as the dataset do not contain newlines that ROUGE-Lsum uses to separate the input with. These scores were then averaged to determine the score of the models. The base model obtained a BLEU score of 0.8099 and ROUGE-L Score of 0.8336 and the finetuned model obtained a BLEU score of 0.8151 and ROUGE-L Score of 0.8396. While the difference between the models are very minimal, this does not completely represent the performance of the models as these metrics are only used to determine if the generated text is close to the reference text, regardless of the context and the overall quality of the generated text. However, it does show that the finetuned model, while not significantly better than the base model, is close to the reference model.

4.2.4 Manual Evaluation Metrics

To determine which of the two models is preferred by Generation Z students at UPV, the researchers conducted a manual evaluation through a survey administered via Google Forms. The survey comprised a total of 93 questions, which were distributed across five separate forms. The first form contained 20 questions, the second 19, the third 20, the fourth 20, and the fifth 14, amounting to 93 questions in total. Each question presented two translation options: one generated by the fine-tuned model and the other by the base model. Respondents were asked to select the translation they preferred in each case. A total of 114 individuals participated in the survey, with 29, 22, 22, 21, and 20 respondents completing Forms 1 through 5, respectively.

The data presented below illustrate respondent preferences between the base and fine-tuned models across the five survey forms, as well as the overall summary of the results. Each graph visualizes the outcomes for an individual form, specifically indicating both the raw number of responses and the corresponding percentages favoring each model. A systematic evaluation for each graph is provided as follows:

Figure 4.5: Form 1 Evaluation

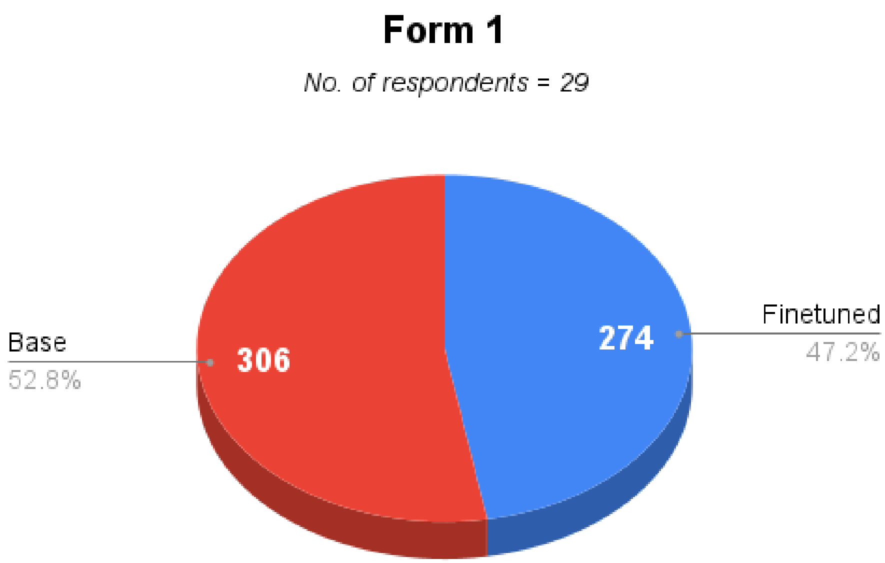


Figure 4.5 shows that among the 29 respondents, 306 responses or 52.8 percent preferred the base model, while 274 responses or 47.2 percent favored the fine-tuned model. This indicates a slight preference for the base model in this particular form. Notably, this result deviates from the overall trend observed in the other four forms, where the fine-tuned model tends to be favored. Form 1 is the only instance in which the base model outperformed the fine-tuned model, suggesting that specific characteristics of this form may have influenced the preferences of the respondents.

Figure 4.6: Form 2 Evaluation

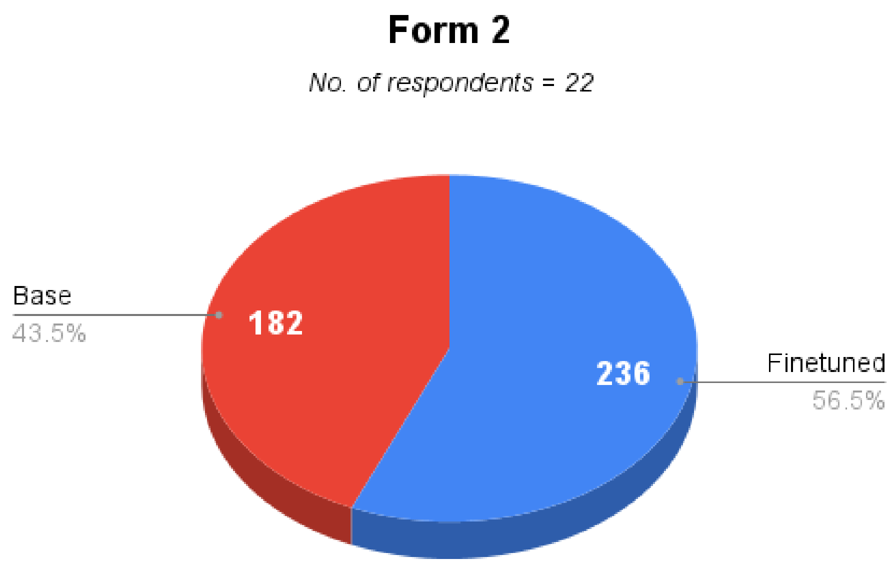


Figure 4.6 implies that among 22 respondents, 236 responses, or 56.5 percent, favored the fine-tuned model, while 182 responses, or 43.5 percent, preferred the base model. This 13 percent margin reflects the clear preference for the fine-tuned model, which is consistent with the overall trend observed across the other forms.

Figure 4.7: Form 3 Evaluation

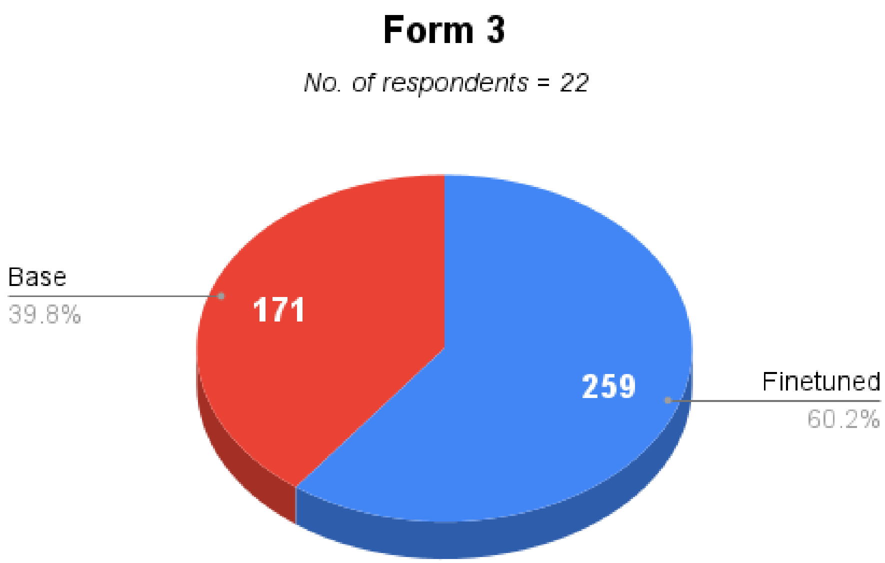


Figure 4.7 illustrates that among the 22 respondents, the fine-tuned model received a significantly higher preference, with 259 responses or 60.2 percent, compared to the base model with 171 responses or 29.8 percent. This 20.4 percent margin represents the widest gap among all forms. This strongly indicates the superior performance of the fine-tuned model on translating, presented in Form 3.

Figure 4.8: Form 4 Evaluation

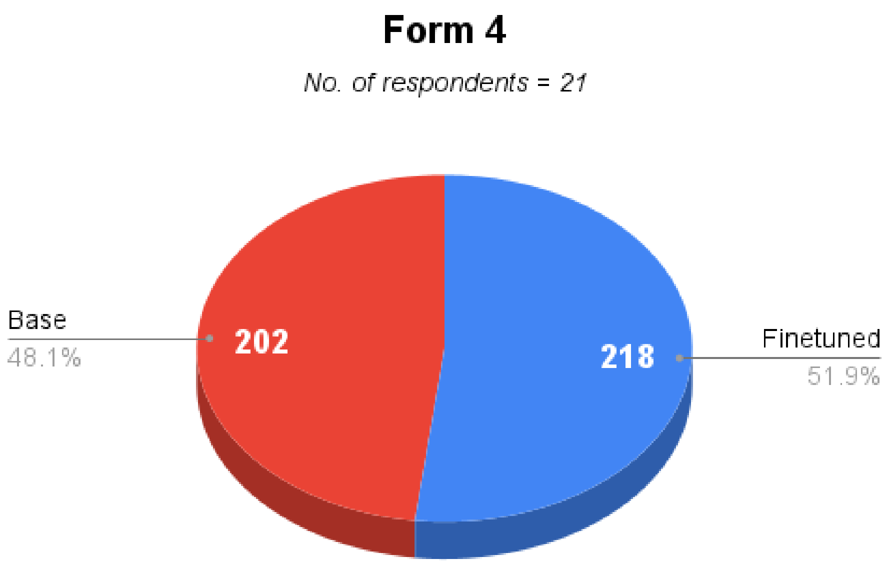


Figure 4.8 highlights that the 21 respondents in Form 4 yielded a nearly even distribution of preferences, with 218 responses or 51.9 percent favoring the fined-tuned model and 202 responses or 48.1 percent preferring the base model. This narrow 3.8 percent difference suggests a comparable level of performance between the two models in this particular form.

Figure 4.9: Form 5 Evaluation

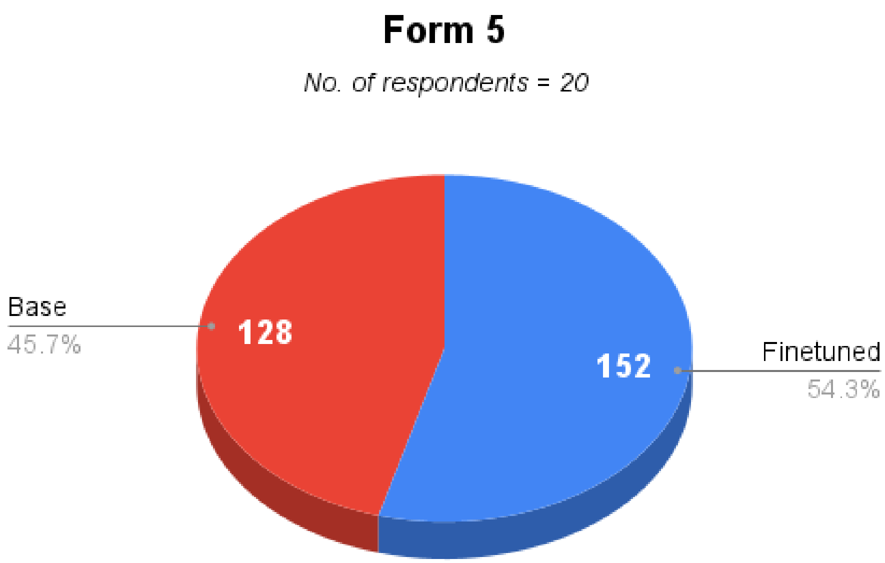


Figure 4.9 conveys that among the 20 respondents in Form 5, 152 responses or 54.3 percent selected the fine-tuned model, while 128 responses or 45.7 percent chose the base model. This 8.6 percent margin reinforces the general trend toward the fine-tuned model across all forms.

Figure 4.10: Summary Evaluation

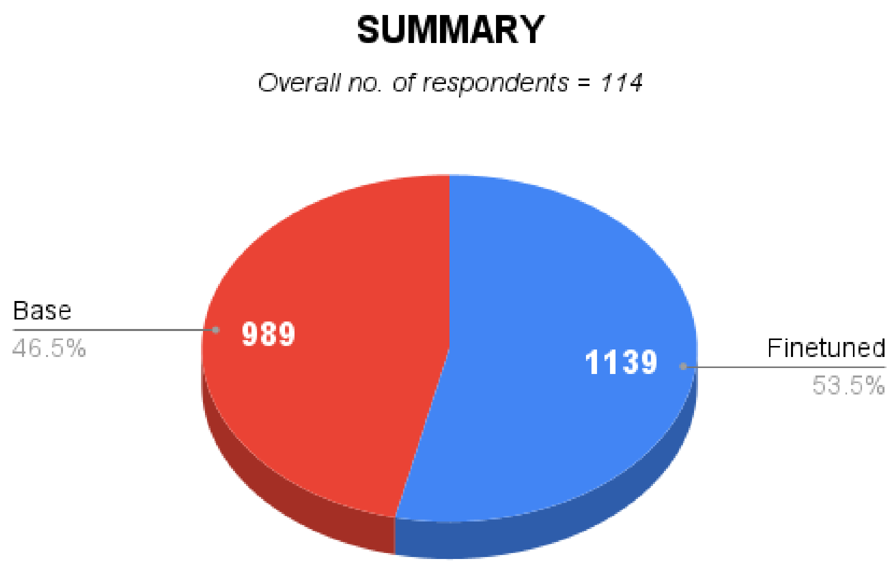


Figure 4.10 presents the overall summary across all five forms, with a total of 114 respondents participating in the survey. In total, the fine-tuned model received 1,139 preferences or 53.5 percent, while the base model garnered 989 preferences or 46.5 percent. The resulting 7 percent margin between the two model indicates a moderate overall preference among Gen Z students at UPV for the fine-tuned model, suggesting its relatively better performance in meeting the participants' expectations for translation quality.

4.3 Summary

The chapter presented the evaluation results and discussions on the performance of the fine-tuned language model for translating Gen Z internet slang into their formal translations. The dataset used for training consisted of 1,703 sentence pairs, combining original and publicly available data. The model was trained for seven epochs, with early stopping employed to prevent overfitting, which was evident from the divergence between training and validation losses.

Evaluation was conducted using both automatic and manual methods. The automatic evaluation, using BLEU and ROUGE-L metrics, showed marginal improvements in the fine-tuned model compared to the base model, suggesting slightly better alignment with reference translations.

To complement the results of automatic evaluation metrics, a manual evaluation was carried out through online surveys among Generation Z students at UPV. Participants compared translations from both models across five forms. Results showed a moderate overall preference for the fine-tuned model, with 53.5% of responses in its favor. While one form showed a slight preference for the base model, the fine-tuned model was generally preferred in the remaining forms, especially in Form 3 where it showed the largest margin.

In summary, the findings indicate that the fine-tuned model slightly outperformed the base model in terms of automatic metrics and showed a modest but consistent preference among target users, supporting its effectiveness in translating Gen Z slang into more formal language.

Chapter 5

Conclusion

In this study, we constructed dataset, containing 1,703 pairs of Gen Z internet slang sentences and their corresponding formal translations. We fine-tuned a zephyr-7B-Beta model and evaluated its performance against the base model. Model training was stopped early to prevent overfitting, and the best model was selected based on validation performance. Both automatic and manual evaluation methods were employed to assess translation quality. Automatic metrics, using BLEU and ROUGE-L, showed that the fine-tuned model slightly outperformed the base model. Manual evaluation, conducted via online surveys with Generation Z students at UPV, indicated a moderate overall preference for the fine-tuned model, which received 53.5% of the total votes. These results suggest that while the improvement in performance was not drastic, the fine-tuned model better aligned with the expectations and preferences of the target demographic.

5.1 Limitations

Language is dynamic and constantly evolving, making it difficult to establish clear boundaries on when slang terms emerge or fade within a generation. Additionally, the dataset created for this study was relatively small, and the number of evaluators involved was limited. In addition, as stated in Section 3.1.3, the computational constraints posed a challenge—loading a model with 7 billion parameters requires approximately 66 GB of memory, while Google Colab provided 16GB of VRAM which is insufficient for high-capacity models.

5.2 Recommendations

Future researchers are encouraged to expand the vocabulary of slang terms used on the Internet and explore more recent trends, taking into account the dynamic nature of language. It is also recommended that future studies utilize a larger and more diverse dataset to improve the robustness of the findings.

References

Ambarsari, S., Amrullah, A., & Nawawi, N. (2020, Aug). The use of online slang for independent learning in english vocabulary. *Proceedings of the 1st Annual Conference on Education and Social Sciences (ACCESS 2019)*, *465*, 295–297. doi: 10.2991/assehr.k.200827.074

Barseghyan, L. (2014). *On some aspects of internet slang.* Retrieved from `https://api.semanticscholar.org/CorpusID:51730779`

binti Sabri, N. A., bin Hamdan, S., Nadarajan, N.-T. M., & Shing, S. R. (2020, Jun). The usage of english internet slang among malaysians in social media. *Selangor Humaniora Review*, *4*(1), 16–17.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative ai at work* (Tech. Rep.). National Bureau of Economic Research.

Crystal, D., & Robins, R. H. (2024, Oct). *Language.* Encyclopædia Britannica, inc. Retrieved from `https://www.britannica.com/topic/language`

Daniel Han, M. H., & team, U. (2023). *Unsloth.* Retrieved from `http://github.com/unslothai/unsloth`

Dua, A., Jacobson, R., Ellingrud, K., Enomoto, K., Cordina, J., Coe, E. H., & Finneman, B. (2024, Aug). *What is gen z?* McKinsey & Company. Retrieved from `https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-gen-z`

Euchner, J. (2023). Generative ai. *Research-Technology Management*, *66*(3), 71–74.

Fernández-Toro, M. (2016, Jun). *Exploring languages and cultures.* Retrieved from `https://www.open.edu/openlearn/languages/exploring-languages-and-cultures/content-section-3.2`

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). *Generative ai and chatgpt: Applications, challenges, and ai-human collaboration* (Vol. 25) (No. 3). Taylor & Francis.

Ghazali, N. M., & Abdullah, N. N. (2021, Dec). Slang language use in social media among malaysian youths: A sociolinguistic perspective. *International Young Scholars Journal of Languages*, *4*(2), 69. Retrieved from `https://www.iium.edu.my/media/`

77652/Slang%20Language%20Use%20in%20Social%20Media%20Among%20Malaysian%20Youths_A%20Sociolinguistic%20Perspective.pdf

Gonzaga, M. (2025, Feb). *"forda convo ang ferson": Analysis of gen z slang in the lens of batstateu faculty members.* Retrieved from https://www.academia.edu/102575643/_FORDA_CONVO_ANG_FERSON_ANALYSIS_OF_GEN_Z_SLANG_IN_THE_LENS_OF_BATSTATEU_FACULTY_MEMBERS

Heydari, M., Albadvi, A., & Khazeni, M. (2024). Persian slang text conversion to formal and deep learning of persian short texts on social media for sentiment classification. *Journal of Electrical and Computer Engineering Innovations (JECEI)*. Retrieved from https://jecei.sru.ac.ir/article_2172.html doi: 10.22061/jecei.2024.10745.731

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., … Chen, W. (2021). *Lora: Low-rank adaptation of large language models.* Retrieved from https://arxiv.org/abs/2106.09685

Ibrahim, A., & Sharief, B. (2023, 10). Intelligent system to transformer slang words into formal words. *NTU Journal of Engineering and Technology*, *2*. doi: 10.56286/ntujet.v2i2.689

Jeresano, E., & Carretero, M. (2022, Feb). Digital culture and social media slang of gen z. *United International Journal for Research & Technology*, *3*(4), 11–25. doi: http://dx.doi.org/10.1314/RG.2.2.36361.93285

Lin, C.-Y. (2004, Jul). Rouge: A package for automatic evaluation of summaries. *Meeting of the Association for Computational Linguistics*, 74–81.

Liu, J., Zhang, X., & Li, H. (2023, Aug). Analysis of language phenomena in internet slang: A case study of internet dirty language. *Open Access Library Journal*, *10*(08), 1–12. doi: 10.4236/oalib.1110484

Liu, S., Gui, D.-Y., Zuo, Y., & Dai, Y. (2019, Jun). Good slang or bad slang? embedding internet slang in persuasive advertising. *Frontiers in Psychology*, *10*. doi: 10.3389/fpsyg.2019.01251

Mantiri, O. (2010, 03). Factors affecting language change. *http://ssrn.com/abstract=2566128*. doi: 10.2139/ssrn.2566128

Maulidiya, R., Wijaya, S. E., Mauren, C., Adha, T. P., & Pandin, M. G. R. (2021, Dec). *Language development of slang in the younger generation in the digital era.* OSF Preprints. Retrieved from osf.io/xs7kd doi: 10.31219/osf.io/xs7kd

McArthur, T. (2003). *Concise oxford companion to the english language* (1st ed.). Oxford University Press.

Nguyen, T. T., Wilson, C., & Dalins, J. (2023). *Fine-tuning llama 2 large language models for detecting online sexual predatory chats and abusive texts.* Retrieved from https://arxiv.org/abs/2308.14683

Nocon, N., Kho, N. M., & Arroyo, J. (2018, Oct). Building a filipino colloquialism translator using sequence-to-sequence model. *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2199–2204. doi: 10.1109/tencon.2018.8650118

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2001). Bleu: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*. Retrieved from https://dl.acm.org/citation.cfm?id=1073135 doi: https://doi.org/10.3115/1073083.1073135

Suslak, D. F. (2009). The sociolinguistic problem of generations. *Language & Communication*, *29*(3), 199–209. Retrieved from https://www.sciencedirect.com/science/article/pii/S0271530909000196 (Reflecting on language and culture fieldwork in the early 21st century) doi: https://doi.org/10.1016/j.langcom.2009.02.003

Teng, C. E., & Joo, T. M. (2023). Is internet language a destroyer to communication? In X.-S. Yang, R. S. Sherratt, N. Dey, & A. Joshi (Eds.), *Proceedings of eighth international congress on information and communication technology* (pp. 527–536). Singapore: Springer Nature Singapore.

Vacalares, S. T., Salas, A. F. R., Babac, B. J. S., Cagalawan, A. L., & Calimpong, C. D. (2023, Jun). The intelligibility of internet slangs between millennials and gen zers: A comparative study. *International Journal of Science and Research Archive*, *9*(1), 400–409. doi: 10.30574/ijsra.2023.9.1.0456

Vergho, T., Godbout, J.-F., Rabbany, R., & Pelrine, K. (2024). *Comparing gpt-4 and open-source language models in misinformation mitigation.* Retrieved from https://arxiv.org/abs/2401.06920

Zhao, J., Wang, T., Abid, W., Angus, G., Garg, A., Kinnison, J., … Rishi, D. (2024). *Lora land: 310 fine-tuned llms that rival gpt-4, a technical report.* Retrieved from https://arxiv.org/abs/2405.00732