No Cap: Fine Tuning GPT-2 to Translate Slang

Tiffany Mo and Princess Leus

Background

Natural language varies widely across social groups, contexts, and generations. One emerging area of sociolinguistic change is the rapidly evolving Gen-Z slang. Online communities like "Stan Twitter", a community of fans who use the platform to support and discuss pop culture, often use language that diverges significantly from standard formal English. This can create communication gaps between age groups, cultures, and online communities.

To address this gap, we developed a formal-to-Gen-Z slang translator by fine-tuning a pre-trained GPT-2 model on paired formal and slang sentences. This tool can help to bridge communication gaps between novice and platform dominant users as well as help marketers generate engaging, stylistically accurate content for younger demographics.

Related work

Our project focuses on Text Style Transfer (TST), specifically generating informal, slang-based text from formal input. This falls within the growing domain of Informal Language Processing. Traditional NLP systems, often trained on formal corpora, struggle significantly with dynamic characteristics of Gen Z slang. As contemporary digital discourse relies heavily on emojis, abbreviations, and slang, our project contributes to generating this evolving language.[1]

Slang creates significant challenges for traditional Natural Language Processing (NLP) systems due to its dynamic nature. Our approach falls under Formality Style Transfer (FST), which aims to translate between formal and informal registers while preserving core meaning. This TST task is crucial for developing robust NLP applications capable of processing and generating fluent, context-aware digital language.

TST solutions vary in scope. The Hugging Face genz-slang-generator model[2], for instance, uses a fine-tuned Gemma2-2B model to perform creative generation. This model creates new slang terms based on a user-provided context. In contrast, our project addresses a different TST problem: translational style transfer. Our method, similar to the Gemma2 model's training procedure, relies on fine-tuning a pre-trained transformer (GPT-2) on parallel sentence data. This approach is valuable for content creators and platforms aiming to effectively engage with a

---

[1] Frnd, Sateesh. Cleaning Digital Language: Dealing with Emojis, Abbreviations, and Slang in NLP. Medium. Accessed November 2, 2025. https://medium.com/@sateeshfrnd/cleaning-digital-language-dealing-with-emojis-abbreviations-and-slang-in-nlp-d17af56a552c

[2] https://huggingface.co/MLBtrio/genz-slang-generator

youth demographic by ensuring their generated text is fluent and stylistically accurate, rather than inventing new terms.

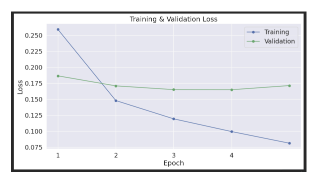
Dataset

The dataset contains 1000 everyday English sentences paired with Gen-Z slang equivalent sentences. The dataset was found on Hugging Face and generated by OpenAI GPT-4.1 Nano. We originally wanted to use a more current and diverse Gen-Z slang dataset[3], but it lacked the formal-slang sentence equivalent needed for our translation task.

Model Implementation

Our model finetunes GPT-2 using Hugging Face's Transformers library and PyTorch to translate formal sentences to Gen-Z slang equivalent sentences. After finding and loading our dataset, we structured our input sequences/training examples to enable the model to recognize special tokens and associate "Formal" and "Slang" with a specific text style. We did this by prefixing all formal sentences from our dataset with the special token <Formal:> and all slang sentences with the special token <Slang:>. We separated each sentence with <Translate> to act as our unique separator token that the model learns to associate with the 'translation' operation. We then introduced the special tokens to the tokenizer's vocabulary and resized the model embeddings. During generation, we prompted the model with this structure: "<|startoftext|> <Formal:> {formal_sentence} <Translate> <Slang:>". The model, having learned this pattern from the training data, will then attempt to complete the sentence sequence by generating the associated slang version of the given input sentence.

Training Parameters

Iteration	**Epochs**	**Learning rate**	**Average training loss**	**Validation loss**	**Perplexity**	**Graph**
1	5	5e-4	0.08	0.17	1.187017	

Settled on original training parameters because loss values were very low and generated results that seemed accurate.

Generation Parameters

[3] https://huggingface.co/datasets/MLBtrio/genz-slang-dataset

Iteration	max_length	top_k	top_p	temperature	Formal sentence	Generated slang translation	Commentary
1	200	50	.95	1.2	I'm really hungry. Do you want to eat lunch with me?	I'm so hungry, let's slide and grab some lunch tmrw, yolo? Do you wanna catch me up? ✌️ I'm super hungry, let's slide and get some grub on our lunch sesh!	High temperature and max length led to high diversity but low coherence. “Tmrw, yolo” are irrelevant to the formal sentence.
2	50	25	2	.8	I'm really hungry. Do you want to eat lunch with me?	No cap, you tryna slide and grab some grub? Yo, I'm totally hungry, can you give me a hand with your sesh, I’m down to chillin' with my friends for a bit? �	Low top_k, top_p, and max_length produced a highly accurate translation, but still outputted some filler phrases (“can you give me a hand with your sesh”)
3	50	75	1.3	.85	I'm really hungry. Do you want to eat lunch with me?	No cap, you tryna slide and grab some grub? Pulling a quick munchies run through my lunch	Used concise, high frequency slang. 0.85 temperature was a sweet spot for coherent slang tokens.
4	50	3	4	1.2	I'm really hungry. Do you want to eat lunch with me?	Yo, you tryna hit the cafeteria?	Concise translation. Low top_k and high temperature resulted in an abrupt, less common translation. It reads well, however, is not as expressive. This highlights the need for mid-range sampling.

For this particular phrase, we found that Iteration 3 had the most effective hyperparameters. Constraining the max_length to be minimally longer than the expected output was an effective evaluation strategy. The nature of slang is typically shorter and abbreviated compared to formal text, so allowing the model to generate excessively long sentences (like in Iteration 1) leads to low-quality, incoherent output.

There were some trade offs in the sample parameters. High temperature showed that the randomness led to irrelevant phrases (ex. “Lunch with me” → “catch me up.”) A mid/high range top_p (ex. 0.95) with a mild temperature allowed the model to select the most semantically relevant tokens, ensuring both fluency and stylistic adherence. Iterations 2 and 3 demonstrate the model’s ability to correctly identify key components for replacement:

Formal Sentence Snippet	**Translated Snippet to Slang**
“I’m really hungry”	“No cap, I’m totally hungry” or “I’m mad hungry.”
Do you want to eat lunch with me?"	"you tryna slide and grab some grub?"

Results

Epoch	**Training Loss**	**Validation Loss**	**Validation Perplexity**
1	1.1265	0.1979	1.2188
2	0.1545	0.1720	1.1876
3	0.1224	0.1673	1.1822
4	0.1015	0.1645	1.1788
5	0.0832	0.1698	1.1851

The model’s performance was monitored across five epochs using loss metrics. Tracking the Training Loss and Validation Loss throughout helped indicate model convergence and detecting overfitting. The sharp decrease in Training Loss (from 1.13 to 0.08) confirms the model learned the patterns present in the training set. The Validation Loss reached its minimum on the 4th Epoch before slightly increasing to 0.168 in the 5th Epoch. This suggests that Epoch 5 marks the onset of overfitting, where the model began memorizing training details at the expense of generalization.

#	**Formal sentence**	**Generated Slang Translation**	**Commentary**
1	This jacket looks very good on you.	This jacket's fire on you, no cap! 🔥	• Concise • "very good" → "fire" • “No cap” used in the right context • Fire emoji

2	Did you hear about the drama that happened at the concert last night?	Yo, did ya catch the drama that just dropped at the concert yesterday? It's lit! Yo, peep the vibes	• Overgeneration • Correct tone “yo”, “catch the drama”, “dropped”, “lit” • Repeats “yo”, irrelevant “peep the vibes”
3	I'm going to fail this exam if I don't study	No cap, I'm down to fail this test if I don't get enough sleep LOL I'm just tryna chill at home and binge	• Wrong semantic • Model ignores input condition (“if I don’t study”) but model introduces an irrelevant condition (“if I don’t sleep”) • Irrelevant “tryna chill”
4	I'm shocked how funny this actor is.	That actor is totally hilarious! That flick is straight fire, no cap, no cap!	• 2nd sentence comments on the “flick” rather than the actor • Model thinks “straight fire” is synonymous with “funny”
5	When should we schedule our meeting	Yo, let’s link up	• Concise
6	I want to relax once I get home	I might chill at my place later, fr fr fr fr.	• Accurate translation of “relax”, “home” → “chill at my place” • Redundant “fr”
7	We should try and go to an event this Friday.	Wanna hit up a good event this weekend? It's gonna be a vibe! 🔥�	• Adds a positive, idiomatic conclusion “gonna be a vibe”

These tests demonstrate strong Style Transfer Strength and fluency when translating formal sentences to slang. Common traits of successful translations include concision (ex. “Schedule our meeting” to “link up”) and precise lexical substitution (ex. “good on you” to “fire, no cap”). The model

Discussion

Despite achieving low training loss, several qualitative results presented notable, unexpected findings related to inference control and semantic coherence15. Under extreme decoding parameters (high temperature, minimal top_k), the model became trapped in repetitive token sequences (e.g., “frfrfrfr” and “Let's fucc it a hand with my hw”). These surprising failures underscore the critical necessity of strict inference constraints to prevent the model from exploiting the high randomness of certain decoding parameters.

Limitations

We noticed that the dataset was limited in the topics it covered and mainly covered different ways to say hanging out, eating, sleeping, and/or being tired, happy, or excited. Because of this, our model tended to generate better slang translations for formal sentences that covered these topics. The model accurately utilizes slang and emojis as punctuation (ex. no cap, fr, 🔥), successfully mimicking the structure of Gen Z discourse. However, two significant limitations persist: semantic drift and over-generation. The model occasionally failed at content preservation, instead generating a highly fluent but contextually incorrect response (ex. Confusion “study” with “sleep” or commenting on the “flick” instead of the “actor”). Furthermore, the autoregressive nature leads to redundant text and excessive repetition of affirmation tokens (ex. “fr fr fr fr”). While the model has mastered the tone, it still struggles with achieving reliable semantic constraint remains the primary challenge.

Future Research and Practical Implications

Further research on translational style transfer models should investigate evolving and regional slang. These advancements would enhance future models’ utility for stylistically authentic content generation. This project can be extended for multiple practical applications. For example, supporting communication tools that adapt messaging for different audiences to help marketers and content creators tailor language to younger or older demographics. Since our model can translate bidirectionally, social platforms could use its slang-to-formal translation functionality for content moderation by allowing them to better detect sentiment and toxicity written in informal language that traditional models may misinterpret.

Code Availability

Github Repository