

Writing Style Text Transfer

Exploring tools to maintain writing style in the age of LLMs

Nathan Harris

CSPB 3832 Natural Language Processing - Summer 2025

The Problem

- Homogenization of text online
- Making your writing stand out
- Text Style Transfer in Natural Language Processing.
 - Changing stylistic properties of text while preserving meaning.
 - Examples: Formal -> Informal, Positive -> Negative



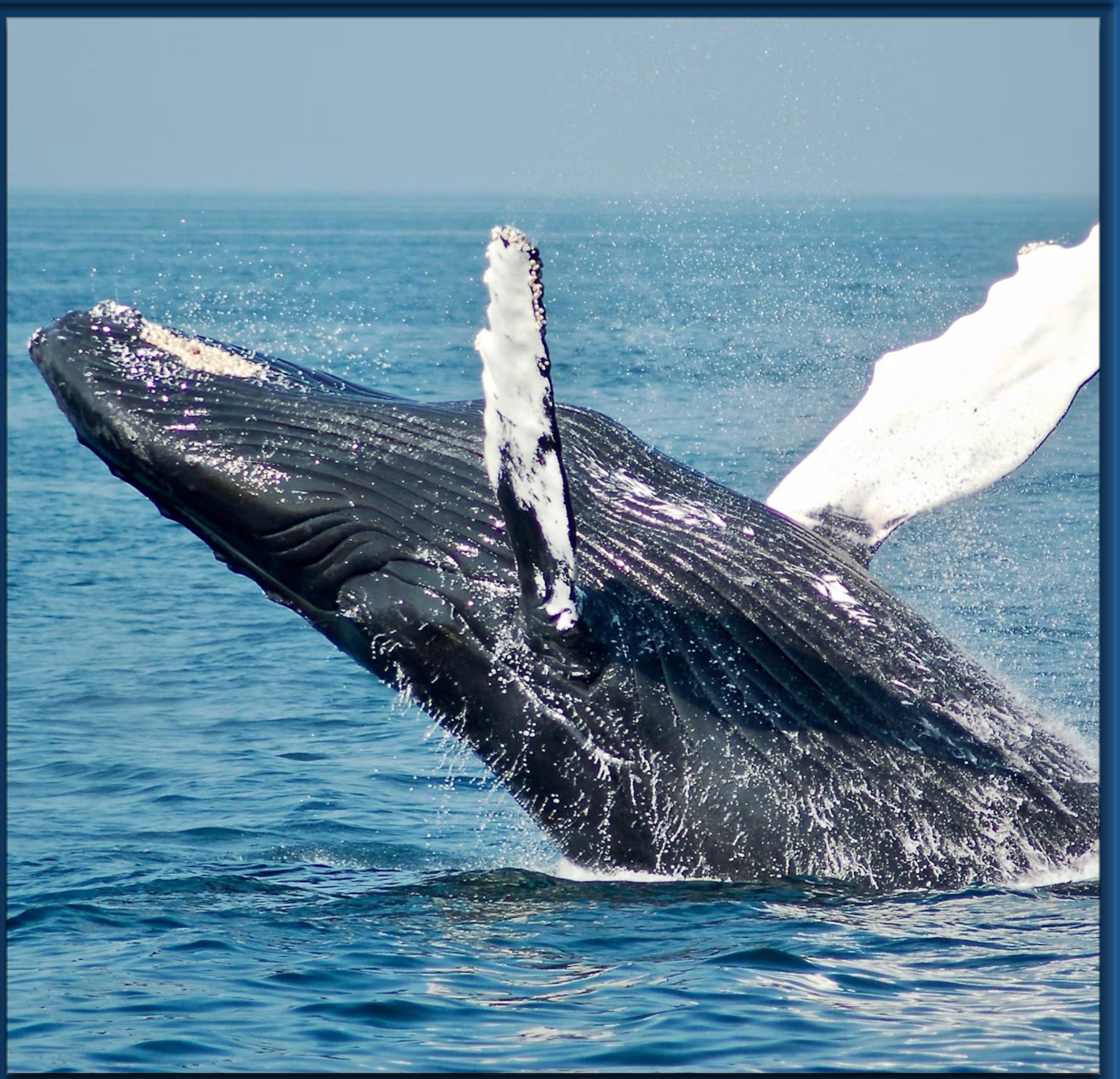
Motivation and Goals



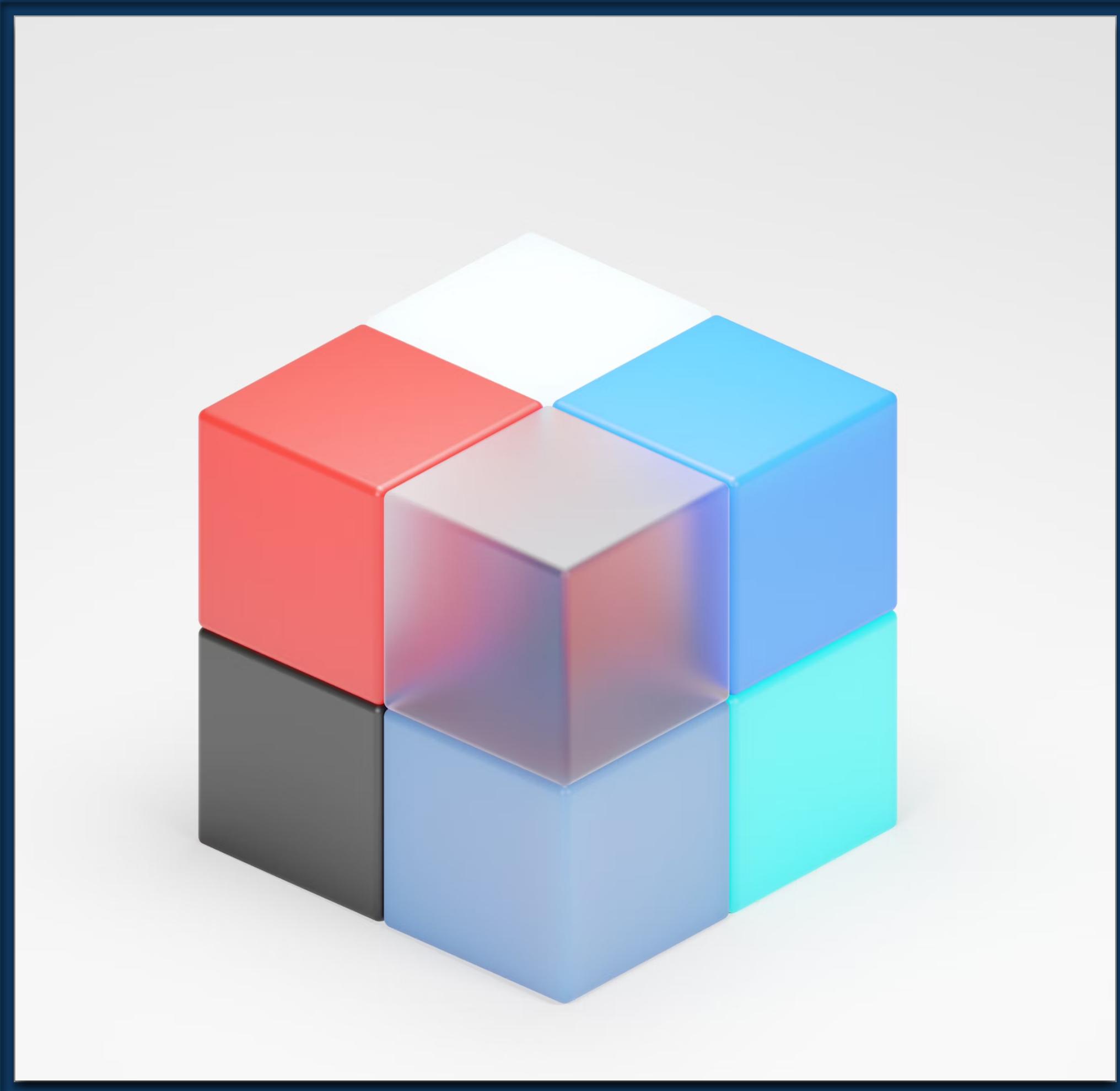
- Journal Writing in my formative years to explore voice
 - Inspired by Cormac McCarthy narrative voice
- Imagined a tool to maintain a distinct writing style
- Experiment with different options
 - Start with fine tuning a pretrained model
 - If time permits, compare with prompting LLMs

Dataset Description

- Author: Herman Melville
 - Author's Text: Moby Dick
 - Neutral Text: Author's Text converted using ChatGPT
 - Length - 100 paragraphs or chunks
- Hugging Face Formatted Dataset
 - Author
 - Input (Neutral Style Text)
 - Target (Stylized Text)



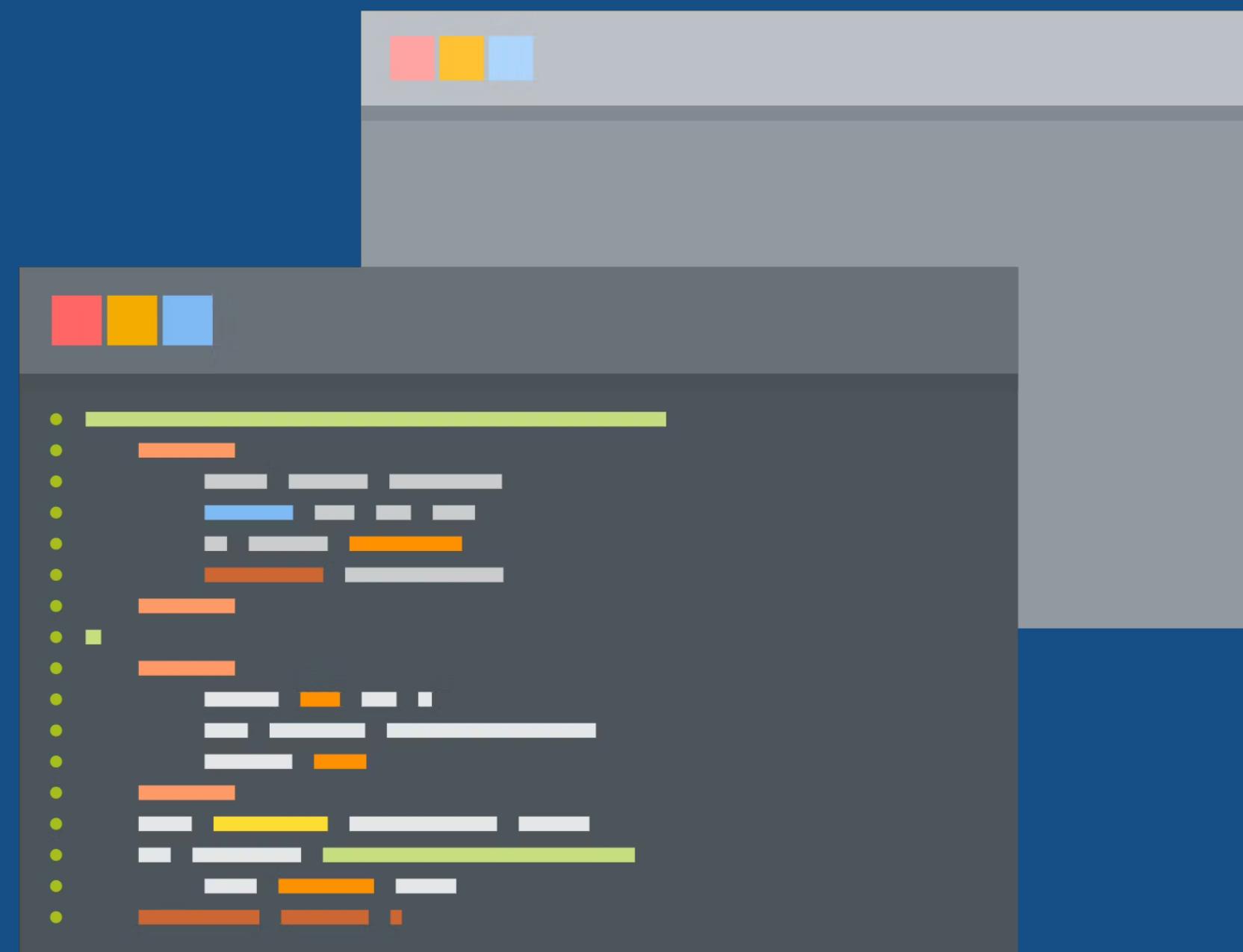
Approach Overview



- Fine tuning a pretrained model
 - Bias a model towards a larger dataset
 - context window of an LLM is limited size
- Use the Model for a niche task, such as Text Style Transfer
- Focused on privacy and open source tools
- Bleu, Rouge, and Perplexity libraries for evaluation

Tools and Techniques

- Hugging Face Transformers in Google Colab
- Model - GPT2-Medium
- Experimented with other models, Google Flan T5 (small and large), Tinyllama, GPT-2 small
- NLTK, Bleu, Rouge, Perplexity, Claude AI



Experiment and Analysis



- Text generation instead of style Transfer
- Experimented with Automating generating neutral text. GPT2 and Google Flan T5
- Crafted a custom GPT on ChatGPT to convert stylized text in to neutral text with minimal prompting
- Attempted different models for text transfer.
- GPT-2 Medium was the first to generate usable results

Results

- Bleu Score:
 - Good at style changes
- Rouge-1 Score:
 - Original meaning likely lost
- Perplexity Score
 - Fluent



Sample Result

Input: Manhattan is surrounded by docks, similar to how coral reefs surround islands. Business activity covers it like waves. The streets lead toward the water. At the southern tip is the Battery, where the pier is hit by ocean waves and cooled by sea breezes that were far from land not long ago. Many people are gathered there, watching the water.

Generated: As if a great mariner with his ship at anchor near this place might have been foretold about as he was approaching its shores, so on account should we now look into every part or section within our city; for such places may be counted upon to receive these tides, which would render them no longer seaward but rather inland—like those mighty swells they call seas--and thus furnish us

Sample Result

Input: The sky is blue...

Generated: I am a man, and an old friend of mine. You are too young to understand me now that I have written you this letter; but when we met here at last night it was my wish to meet again with one who knows all about myself...

Challenges and Lessons Learned



- Underestimated Complexity of project
- Self imposed constraints made it difficult
- I imagined I would get workable results and explore parameters and different options. Instead I struggled to just get results.
- Results would probably be better with a better dataset - vetted by a human

Conclusion and Future Work

- If I had more time
- Explore Prompt Engineering
- Explore Retrieval Augmented Generation
- Explore compressing writing style instead of Dataset



The End. Thank You

Images provided for free by unsplash.com