

Investigating the Capabilities of ChatGPT in Poetry Generation and Analysis

Elizabeth Gronski
Informatics '24

Team Introduction

During my research capstone project, I collaborated with Professor Melanie Walsh, an Assistant Professor at the Information School, on her research. My responsibilities included collecting data, performing comprehensive textual and classification analyses of ChatGPT results, and contributing towards ideation for future analysis.

Problem Context

The emergence of language models such as ChatGPT has revolutionized the way we interact with technology, including in creative and artistic fields. Producing fiction and poetry with ChatGPT has been highlighted by journalists, by the general public, and even by the company itself. Because poetry is a unique, distinctive form of language, it poses an interesting challenge and test case for language models, and it can potentially help illuminate their broader linguistic capabilities and limitations.

Problem Statement

In this research project, we aim to explore ChatGPT's ability to analyze and categorize poetry, which could have significant implications for understanding AI-powered literary analysis. We hope to contribute to a better understanding of ChatGPT's capabilities and limitations regarding poetry and, in turn, to learn more about how the model is trained and works.

Key Research Insights

We collected a large corpus of poetry data from multiple sources, including the Poetry Foundation and the Norton Anthology. We utilized web-scraping and manual collection to gather the data. We used the ChatGPT API to tag each poem by its poetic form (from a range of provided options), and we then analyzed the model's accuracy in identifying each form.

In analyzing ChatGPT's ability to recognize poetic forms, we found that the model works well with popular poetic forms, such as limericks and ballads, but also, perhaps surprisingly, on little-known forms, such as pantoums and ghazals. However, the model does not perform well with forms like pastorals and terza rimas — for reasons we are still investigating. Additionally, when analyzing the poems generated by ChatGPT, we discovered a disproportionately excessive use of the word "love" and of first plural pronouns, such as "we," "us," and "ours."

These initial findings provide insight into ChatGPT's limitations and assumptions regarding poetry and may help researchers identify similar patterns. We are still trying to determine the extent to which these poems may be memorized by the model.

Ethical Considerations

Many of the poems that we fed into the model and asked the model to identify are currently under copyright. We also prompted the model to produce poems in the style of a handful of authors who are still currently living. These represent ethical complications and quandaries because many artists and writers are angry that their copyrighted works have been used to train large language models without their consent. For these reasons, more than a dozen AI copyright lawsuits are currently ongoing. We are still discussing whether we want to keep the living writers in our ChatGPT-generated poetry corpus and how we want to handle the sharing of this data. Though we are invested in auditing and understanding these technologies, we are fundamentally invested in human artists and authors, and we are still trying to navigate these difficult issues.

Next Steps Beyond Capstone

This project may help bring to light some of the inner workings of ChatGPT. As the research continues, we plan to explore other avenues. Firstly, we will continue to experiment with the prompt provided to the model. Much research has been done into the variability of prompts and how this can significantly change the results provided by the model. We want to experiment with removing most of the poetry text to try to gain more concrete proof of memorization.

Furthermore, we aim to extend this research to other Large Language Models (LLMs). Given ChatGPT's popularity, it serves as a logical starting point. However, comparing as many models as possible would be more beneficial in providing a more accurate depiction of how poetry is incorporated into the AI space. We also plan to expand our corpus of data from the Poetry Foundation poems. Our analysis of the Poetry Foundation poems has

brought us much insight, but we are unaware of who was involved in categorizing these poems in the first place. To diversify our data would give us a more accurate analysis.

Finally, we aim to validate ChatGPT's analysis by comparing it with the expertise of real-life literary experts. Conducting interviews with these experts and recording their categorization of the same poems will provide a valuable benchmark for the model's performance. This comparative analysis will shed light on the model's accuracy and potential benefits for poetry enthusiasts.