

Consistency and Predictability of Short Stories Generated by LLMs

Nicholas Sanaie, Anneliese Brei, Snigdha Chaturvedi



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Abstract

Our research investigates the consistency and predictability of short story generation by Large Language Models (LLMs) by focusing on several narrative elements including character relationships, story emotion arcs and word choices. Our study uses the ROCStories and WritingPrompts datasets alongside short stories generated by GPT-3.5 to explore similarities and predictable patterns in machine generated stories in comparison to human created stories. Analysis of 500 stories from each dataset reveals intriguing patterns, such as less emotional arc variation in machine generated stories and predictable character sentiment evolution throughout the story. There are also slight differences between machine generated datasets. Our research also discusses the challenges of annotating emotional arcs and character sentiments through highlighting the discrepancies between human annotations and those generated by LLMs. One such finding is the preference in machine-generated annotations to confused “Man in a Hole” and “Rags to Riches” emotional arcs. This study contributes to the understanding of the capabilities and limitations of LLMs in narrative generation and also underscores the potential need for new approaches to enhance narrative diversity and creativity. Our research is still ongoing and we are currently collecting and verifying our results so far.

Methods

- We employ several methods to generate, annotate and verify our data/results.
- Generate 500 stories using each dataset.
 - Annotate using GPT-3.5 on emotional arcs, word frequency, character sentiment patterns, showing vs telling, and lexical frequency.
 - Verify annotations using human labeled annotations on a small subset.
 - Compare and contrast against different datasets and between machine and human created stores

Current Findings

- “They found that "Icarus," "Oedipus," and "Man-in-a-hole" were the three most successful emotional arcs.”. This does not match the machine generated results.
- It is difficult to annotate stories accurately using machines.
- There are observable patterns across a large dataset of generated stories
- Machine generated stories typically follow the classic bad to good (roughly) story arc.
- There is a lot of room for improvement!

