UNC DEPARTMENT OF COMPUTER SCIENCE

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

ACL 2026 SAN DIEGO

CASPER in the Machine

Insights into Character Variety In LLM-Generated Stories

Anneliese Brei
UNC Chapel Hill

Abhisheik Sharma
Georgia Institute of Technology

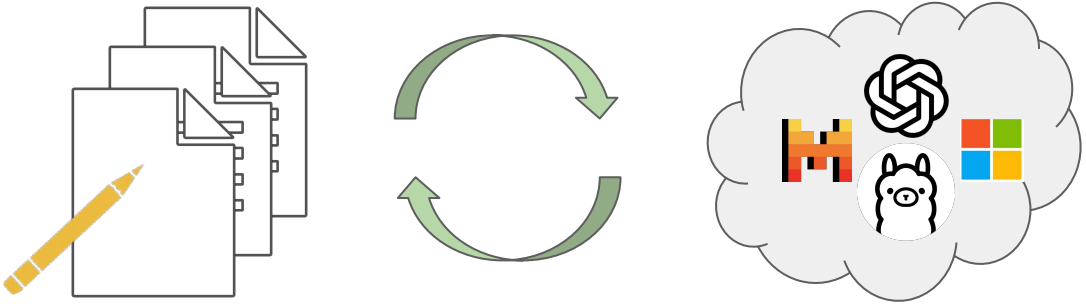
Nicholas Sanaie
UNC Chapel Hill

Lu Wang
University of Michigan

Snigdha Chaturvedi
UNC Chapel Hill

Motivation

Large Language Models (LLMs) used increasingly for process of creative writing



Our Goal

Can we compare characters in human-written stories & characters in LLM-generated stories?

Dimensions of Character Portrayal

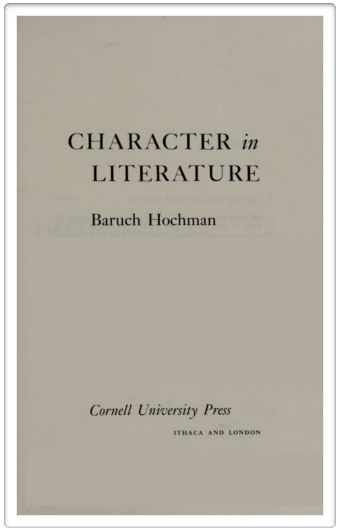
Stylized

Natural

Dumbledore Image: https://www.imdb.com/gallery/rg2936118016/mediaviewer/rm1061797632?ref_=nm_ecw_g_margolis_i
Potter Image: https://www.imdb.com/name/nm0705356/mediaviewer/rm3471437313/

Dimensions of Character Portrayal

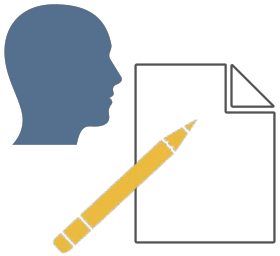
1a.	Stylized
2a.	Coherent
3a.	Whole
4a.	Literal
5a.	Complex
6a.	Transparent
7a.	Dynamic
8a.	Closed



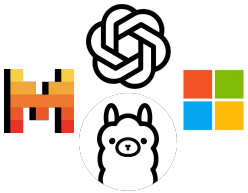
1b.	Natural
2b.	Incoherent
3b.	Fragmented
4b.	Symbolic
5b.	Simple
6b.	Opaque
7b.	Static
8b.	Open

Hochman, Baruch. "Character in literature." *Cornell University Press* (1985).

Corpus Creation

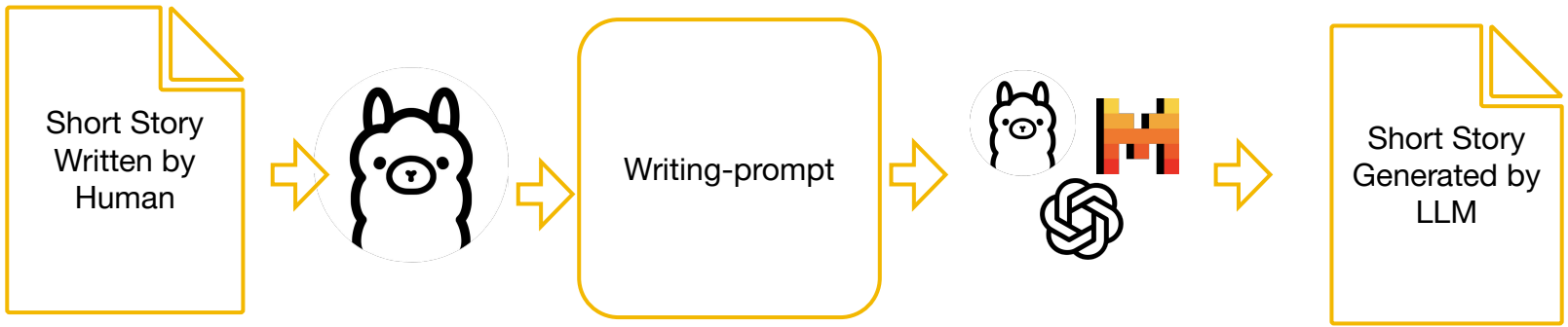


1. Collect 200 human-written from *r/shortstories*
    - Genres: domestic, romance, fantasy, thriller

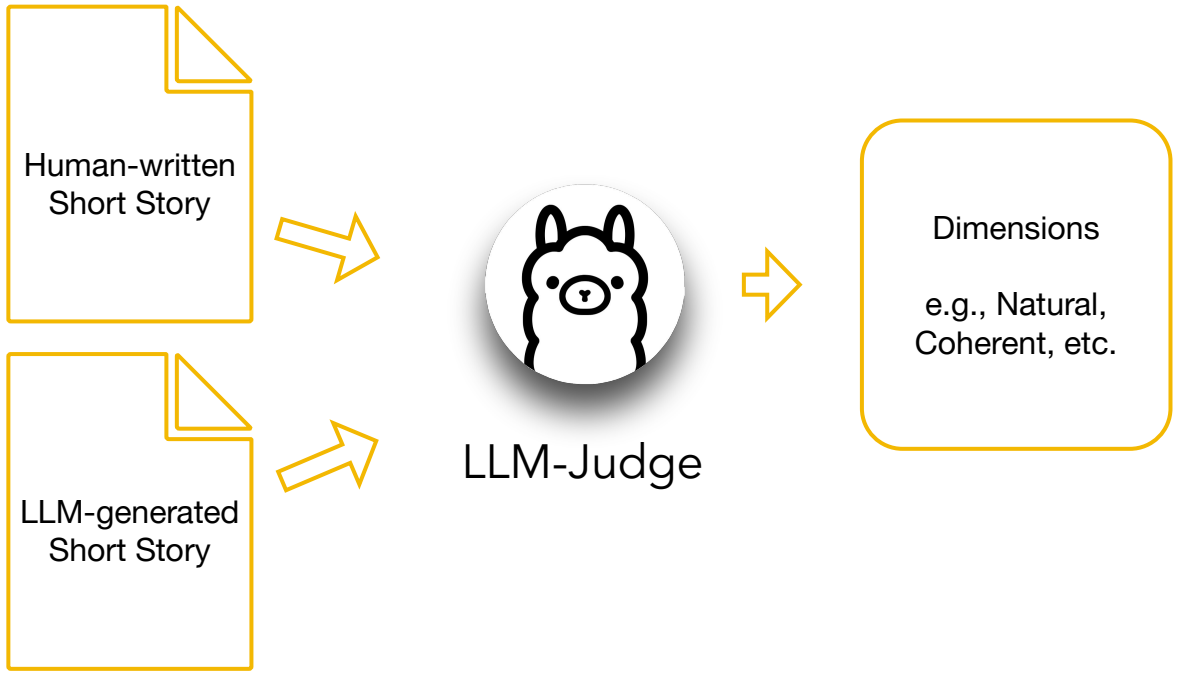


2. Generate 4400 total stories using 8 LLMs from 4 families

CASPER: Data Collection

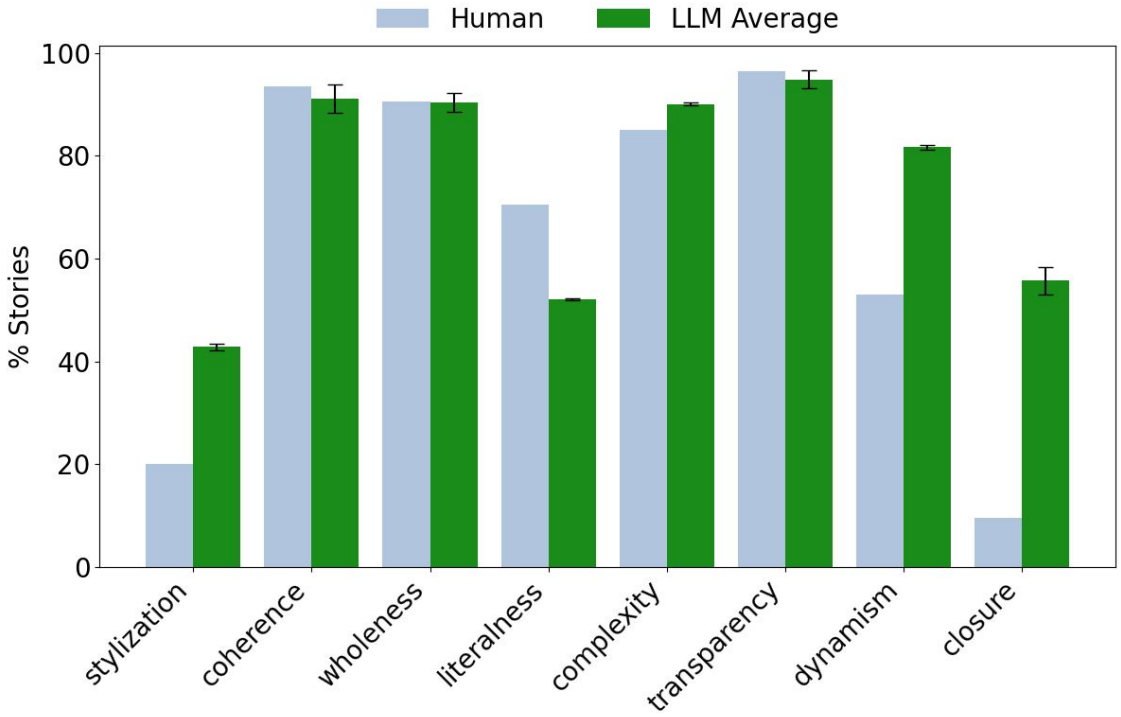


CASPER: Methodology

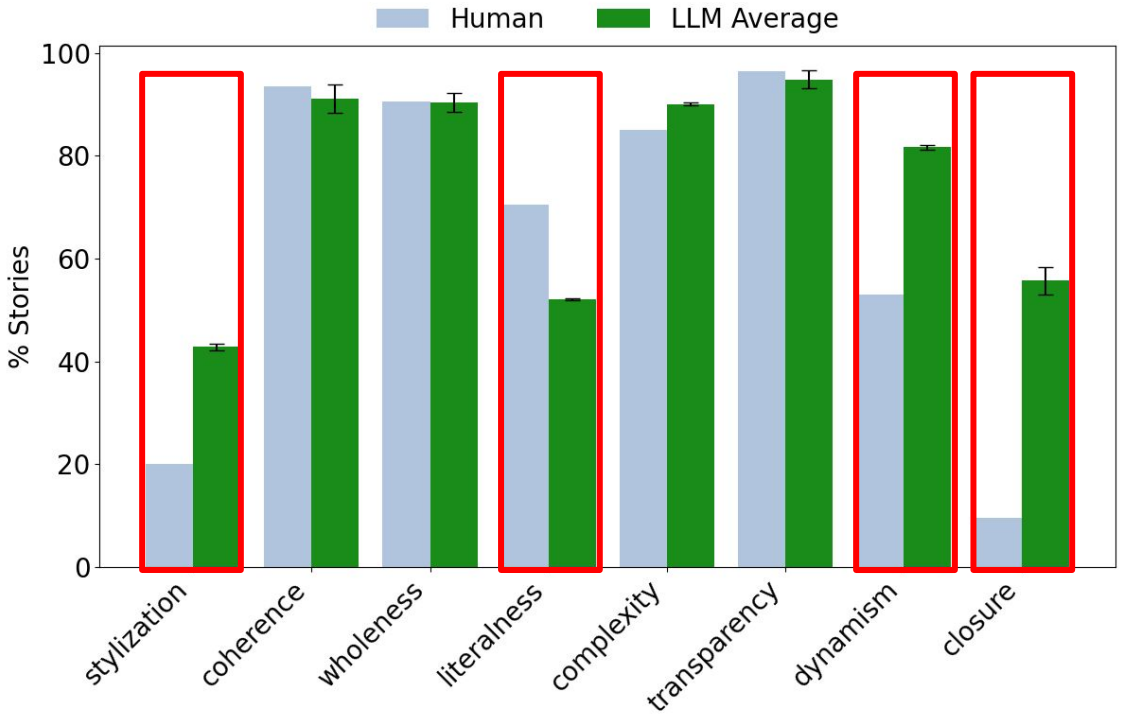


Analysis

How do LLM vs. human-written characters compare?

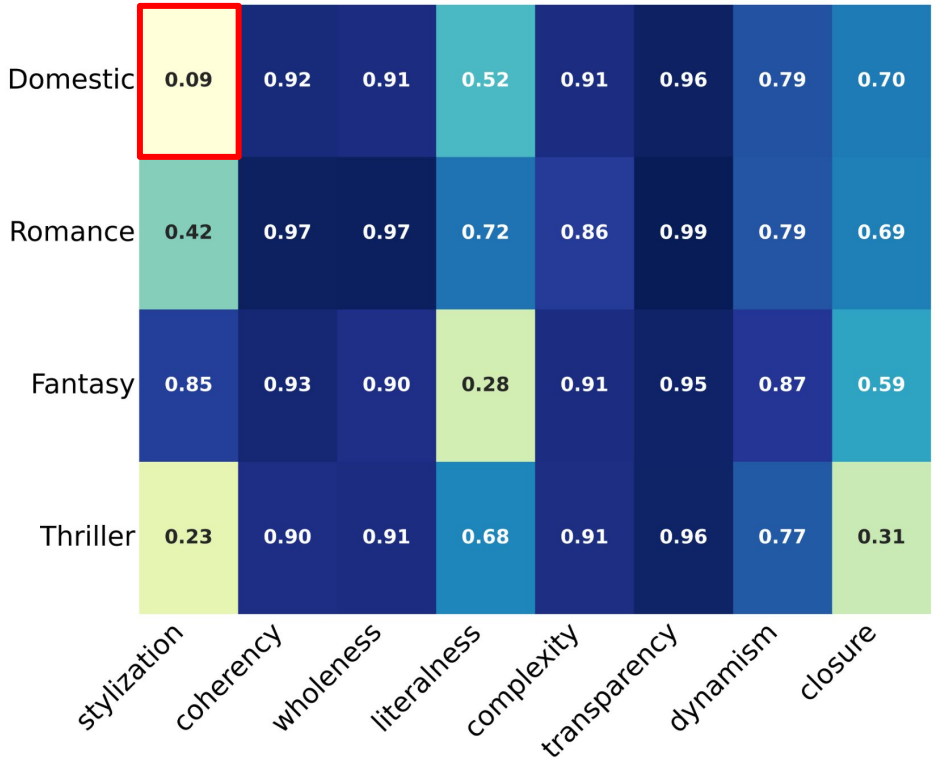


How do LLM vs. human-written characters compare?



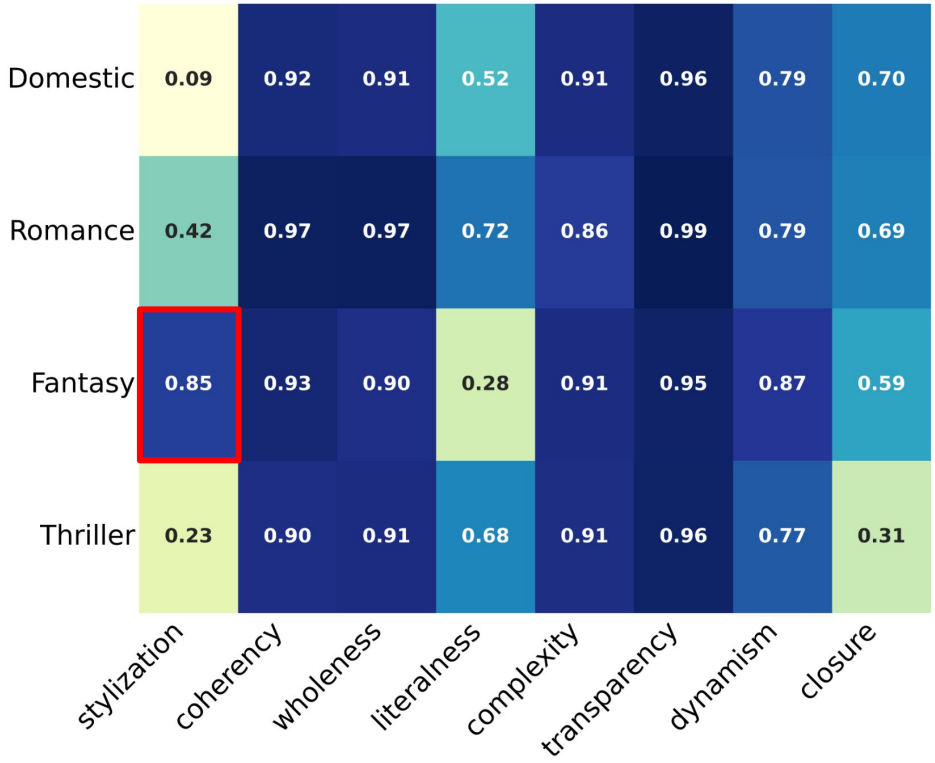
LLM stories “play it safe”

Do LLMs default to particular categories for different genres?



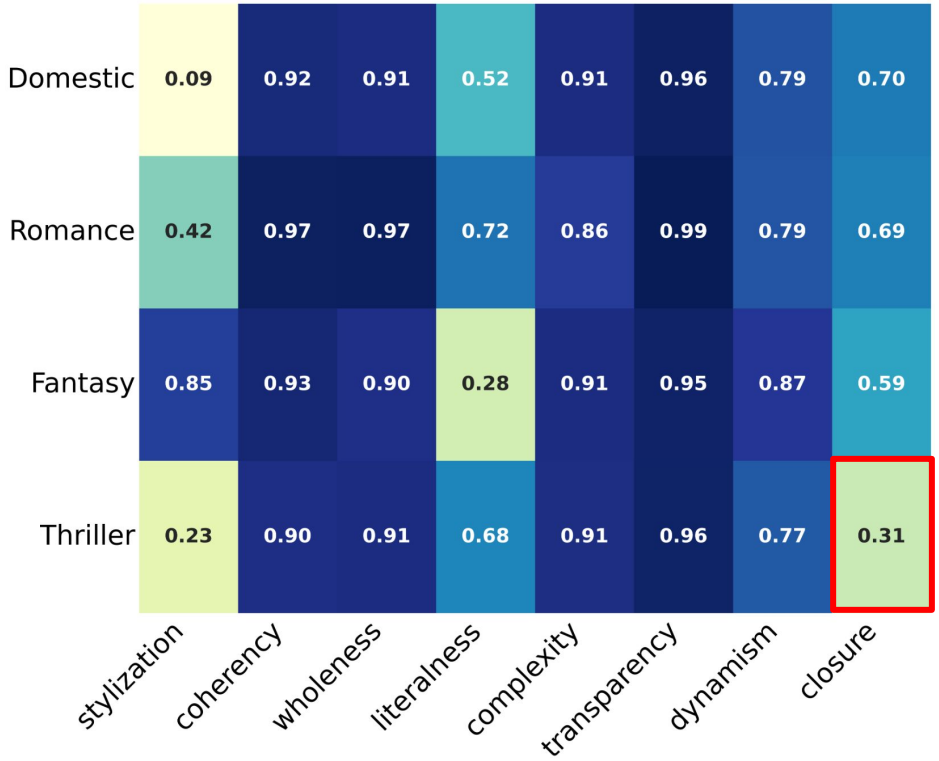
Domestic: ↓stylized

Do LLMs default to particular categories for different genres?



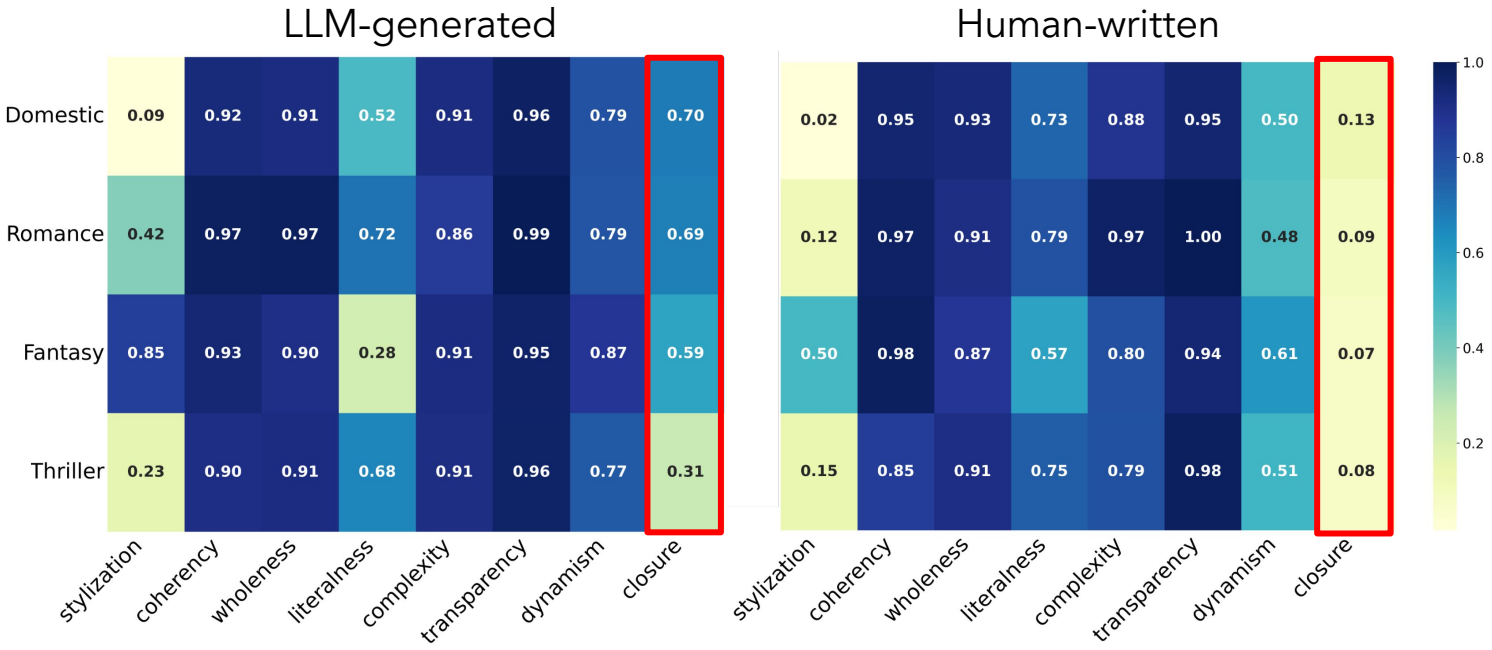
Fantasy: ↑ stylized

Do LLMs default to particular categories for different genres?



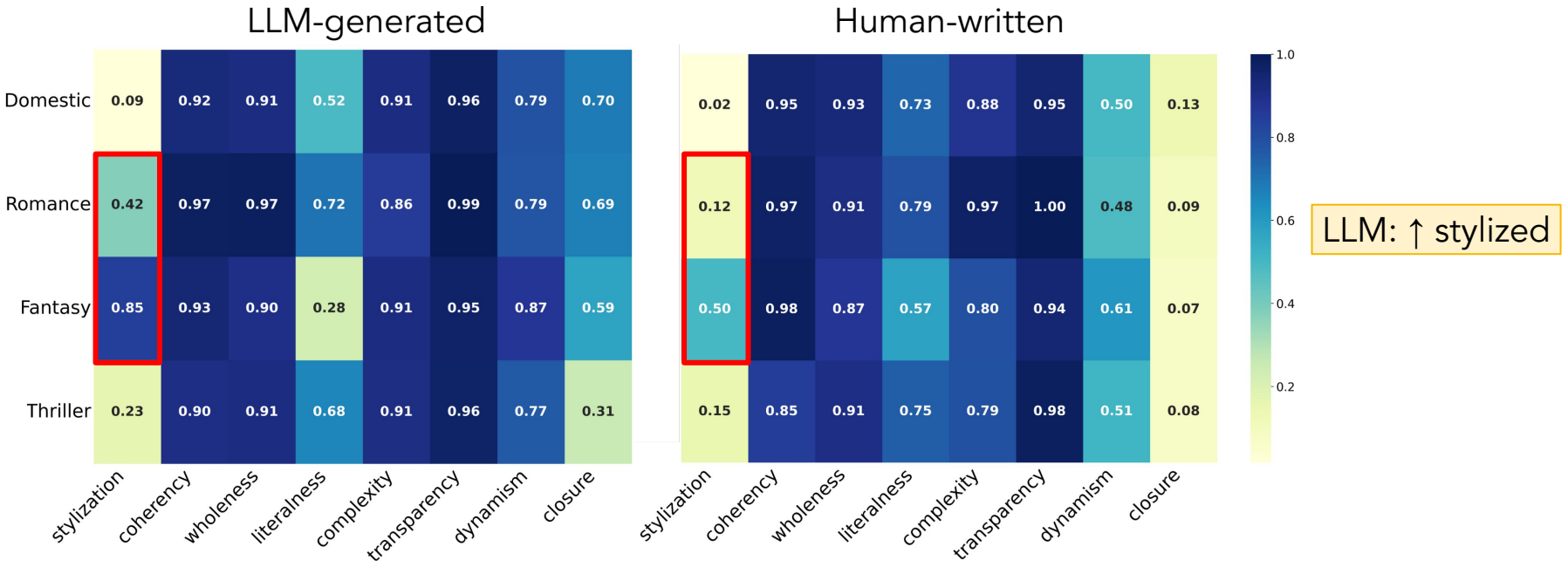
Thriller: ↓ closed

How do LLMs and humans differ across genres?



LLM: ↑ closed

How do LLMs and humans differ across genres?



Other Takeaways

- LLM size does not change types of characters generated.
- Phi family: ↑ diverse
  Llama family: ↓ diverse
- When re-generating stories from a prompt, *literalness* and *closure* show most variability.

CASPER in the Machine: Insights into Character Variety in LLM-Generated Stories

Anneliese Brei[1] **Abhisheik Sharma**[2] **Nicholas Sanaie**[1]
Lu Wang[3] **Snigdha Chaturvedi**[1]
[1]UNC Chapel Hill [2]Georgia Institute of Technology [3]University of Michigan
abrei@cs.unc.edu, asharma914@gatech.edu, nsanaie@unc.edu,
wangluxy@umich.edu, snigdha@cs.unc.edu

Abstract

As LLM-generated text is increasingly used, especially in fictional domains, we explore how much LLM-generated stories differ from human-written stories. In this work, we focus on characters. We borrow definitions from narratology to analyze 8 intricate category-pairs of character, such as *stylization* and *wholeness*. These category-pairs consider more than just basic characteristics. They assess how characters are portrayed within their stories. After automatically inferring categories of characters within both LLM and human-written stories, we compare and contrast these two sets of stories. We consider the following overarching questions: (1) Do LLMs and human-written stories have similar characters? and (2) Do LLMs generate stories with a variety of characters? Our analysis includes research questions that focus on stories generated by popular LLMs and recently published human-written stories. We describe a number of interesting similarities, differences and key takeaways.[1]

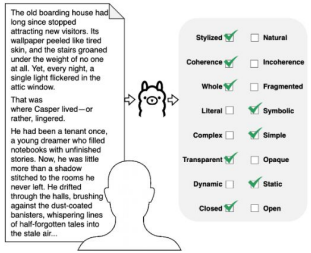


Figure 1: We analyze characters in LLM-generated short stories (left) using 8 category-pairs (right) that consider how characters are portrayed. (e.g., a character might be represented in a *fragmented* way, setting the tone for a disjointed mood within the story). We compare LLM-generated characters with human-written characters other and LLM-generated characters from different model sizes and families.

See more about this work: