AIStorySimilarity: Quantifying Narrative Similarity

Jon Chun

Kenyon College

Introduction

Stories are central to human communication, shaping how we interpret experiences and influence one another through films, media, and narratives. Quantifying the similarity between stories has significant applications, such as detecting intellectual property infringement, identifying hallucinations in AI-generated content, improving search and recommendation engines, and enhancing human-AI collaborations.

Traditional Natural Language Processing (NLP) text similarity metrics are limited to short texts and often result in substantial information loss when applied to longer narratives. AIStorySimilarity leverages LLMs to extract hierarchical story features based on narrative theory to measure novel long-form story similarity which has potential applications in:

❖ Generative AI Intellectual Property Infringement

❖ Search and Recommendation Engines

❖ Human-AI Collaborative Story Writing

Methods

Sim. ref	Genre	Name	Year	Rank
	Adventure	Raiders of the Lost Ark	1981	-
1	Sequel #1	Indiana Jones and the Temple of Doom	1984	2
2	Sequel #2	Indiana Jone and the Last Crusade	1989	2
3	Adventure	National Treasure	2004	10
4	Adventure	Laura Croft Tomb Raider	2001	14
5	Adventure	The Mummy	1999	8
6	Romantic Drama	Titanic	1997	7
7	Black Comedy	Office Space	1999	133
8	Musical	La La Land	2016	83

Table 1: Films similar to Raiders of the Lost Ark
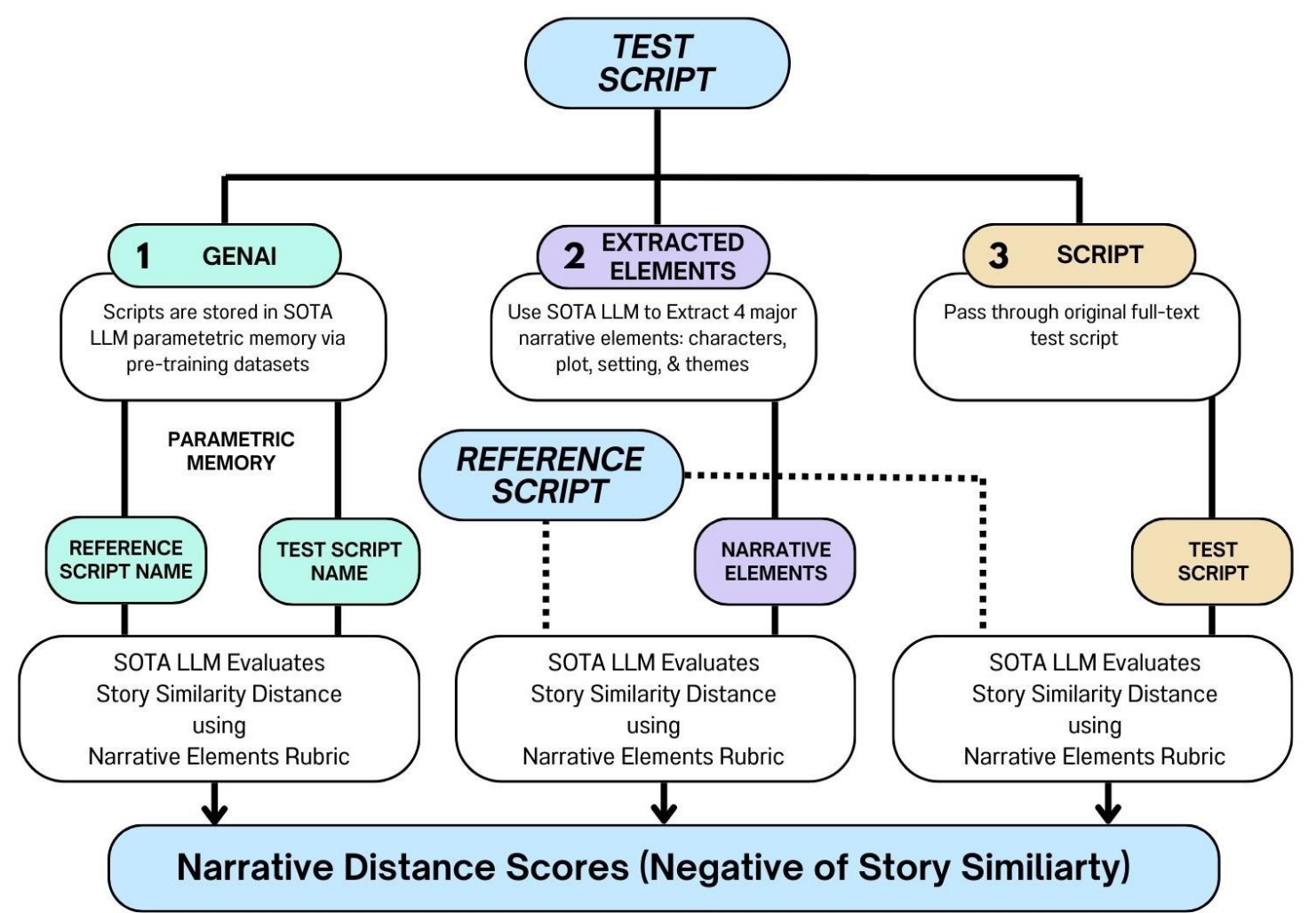
###ELEMENT_FEATURES
Name: Full name of character
Role: Clarifies the character's function within the story, whether they are driving the action, supporting the protagonist, or creating obstacles.
Backstory: This attribute helps to understand the formative experiences that shaped each character, providing insights into their motivations and behaviors.
Strengths: Highlights unique abilities and proficiencies, distinguishing characters by their specific talents and expertise.
Weaknesses: Humanizes characters by revealing vulnerabilities and personal challenges, making them more relatable and multi-dimensional.
Psychology: Uses personality assessments, such as the Big 5 OCEAN (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism) model, to offer deeper insight into character traits.
Beliefs: Offers a window into the ethical and moral framework guiding each character's decisions, crucial for understanding their actions in moral dilemmas.
Motivations: Describes what drives the character to act, including desires, fears, and goals.
SocialDynamics: Explores the nature of interactions between characters, which can be pivotal in character development and plot progression.
Arc: Summarizes how the character changes or grows for better or worse over the story in response to events, decisions, and actions taken
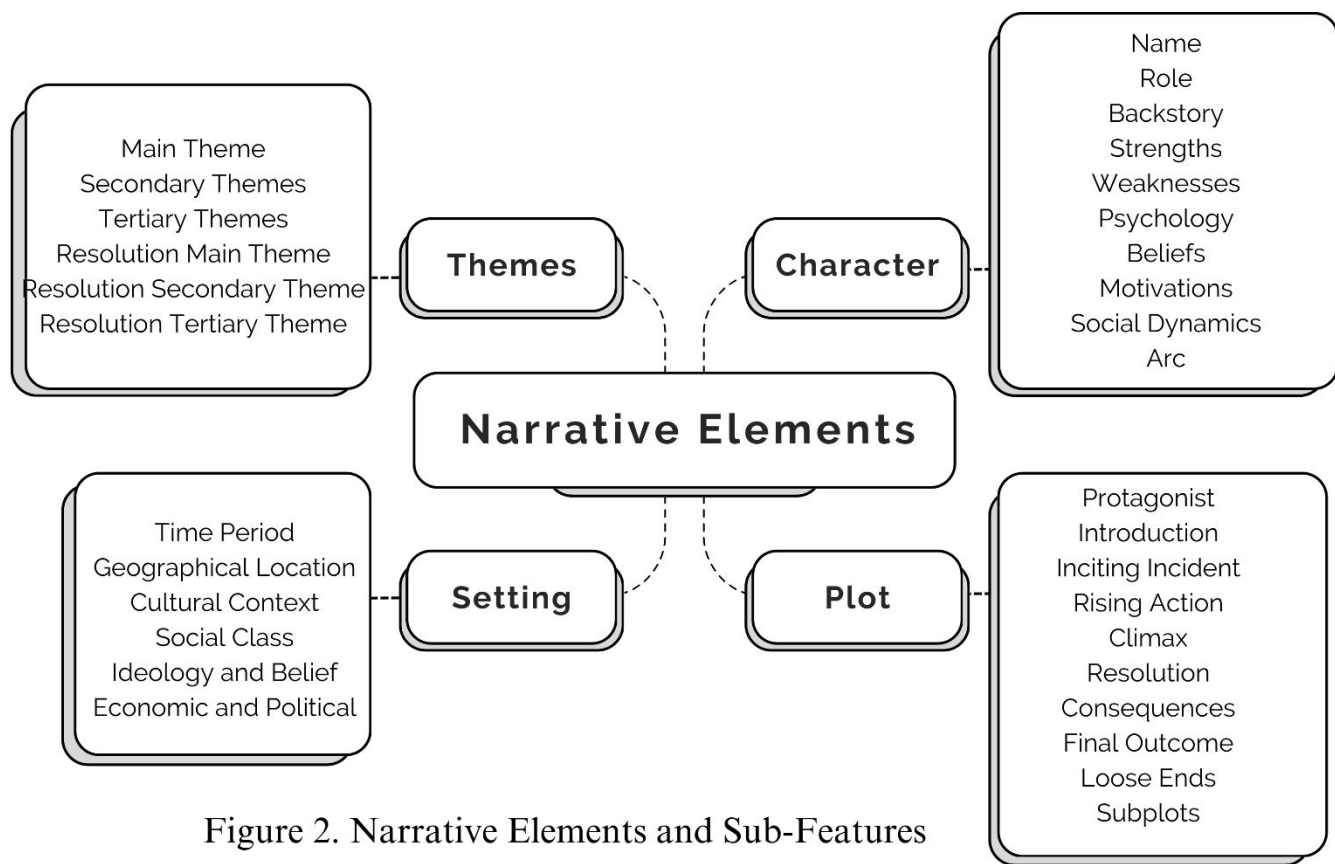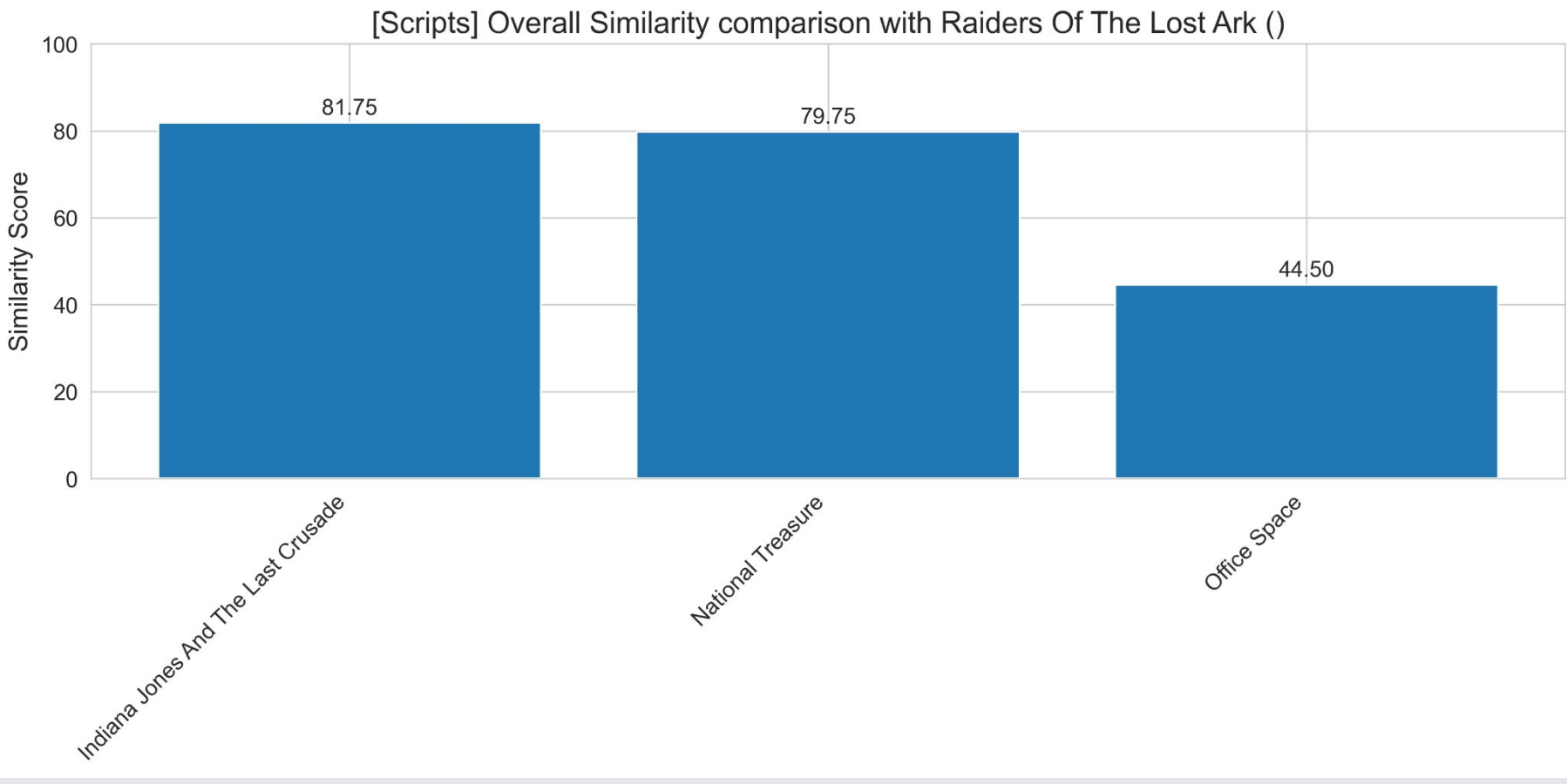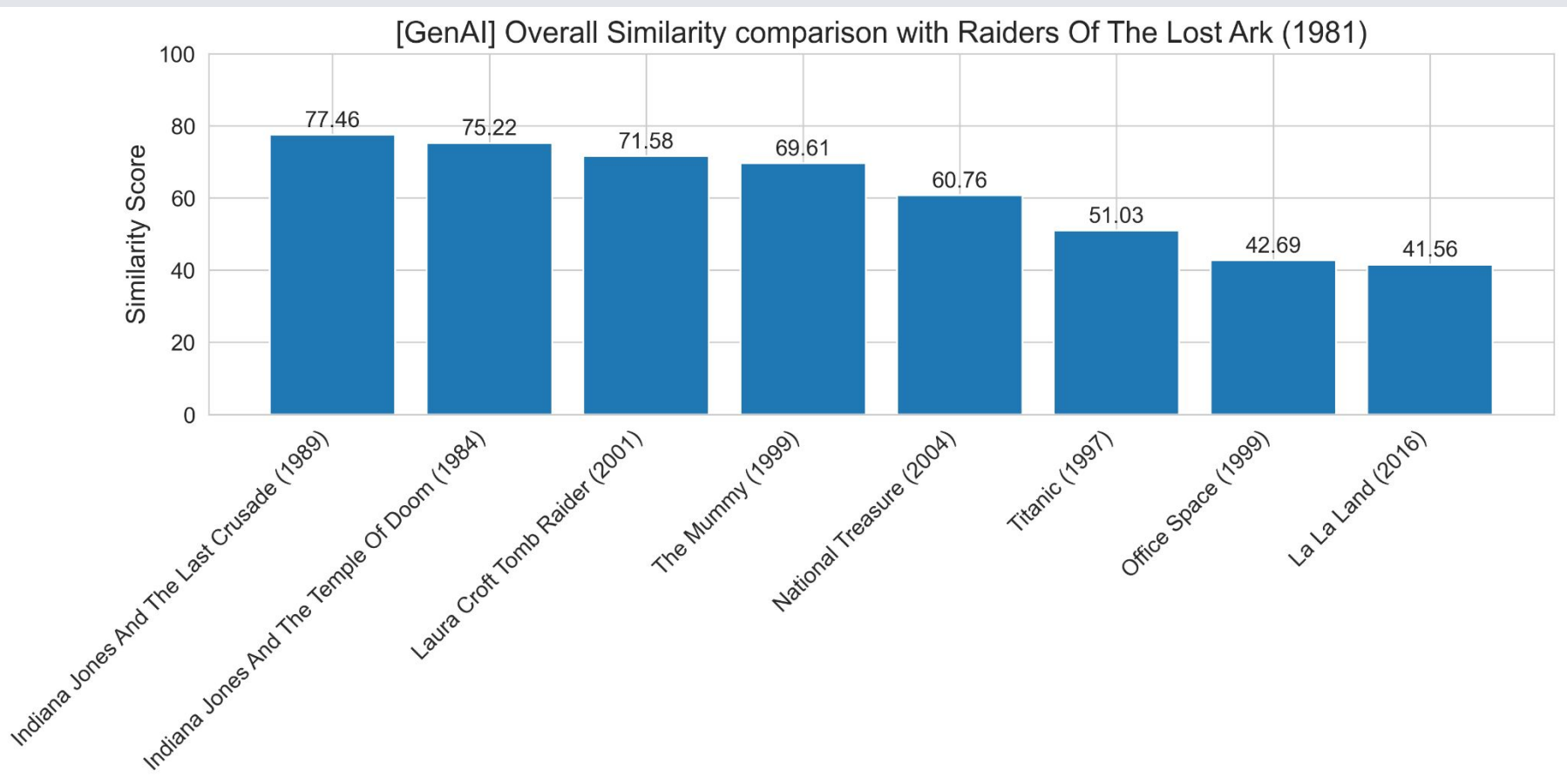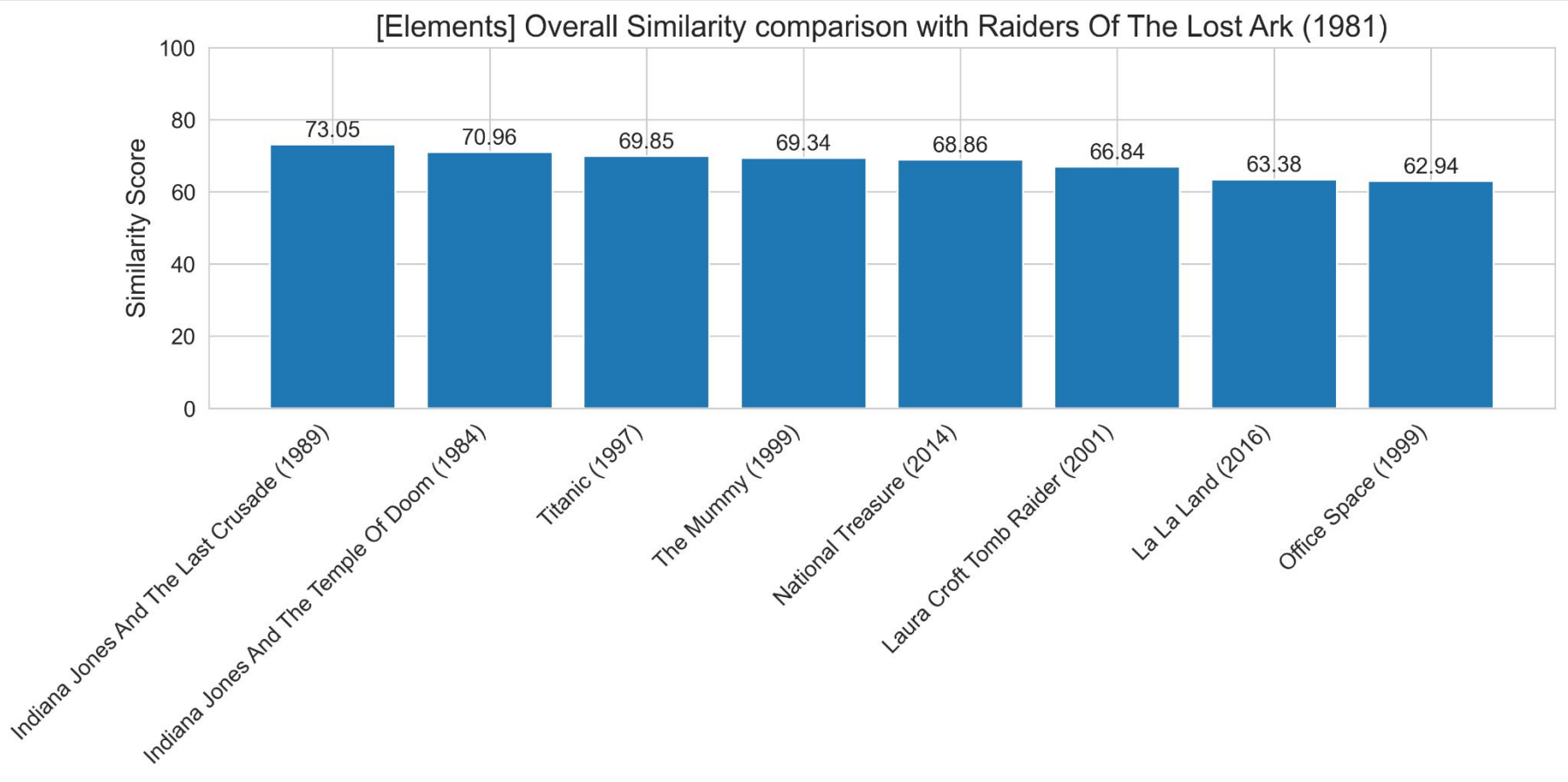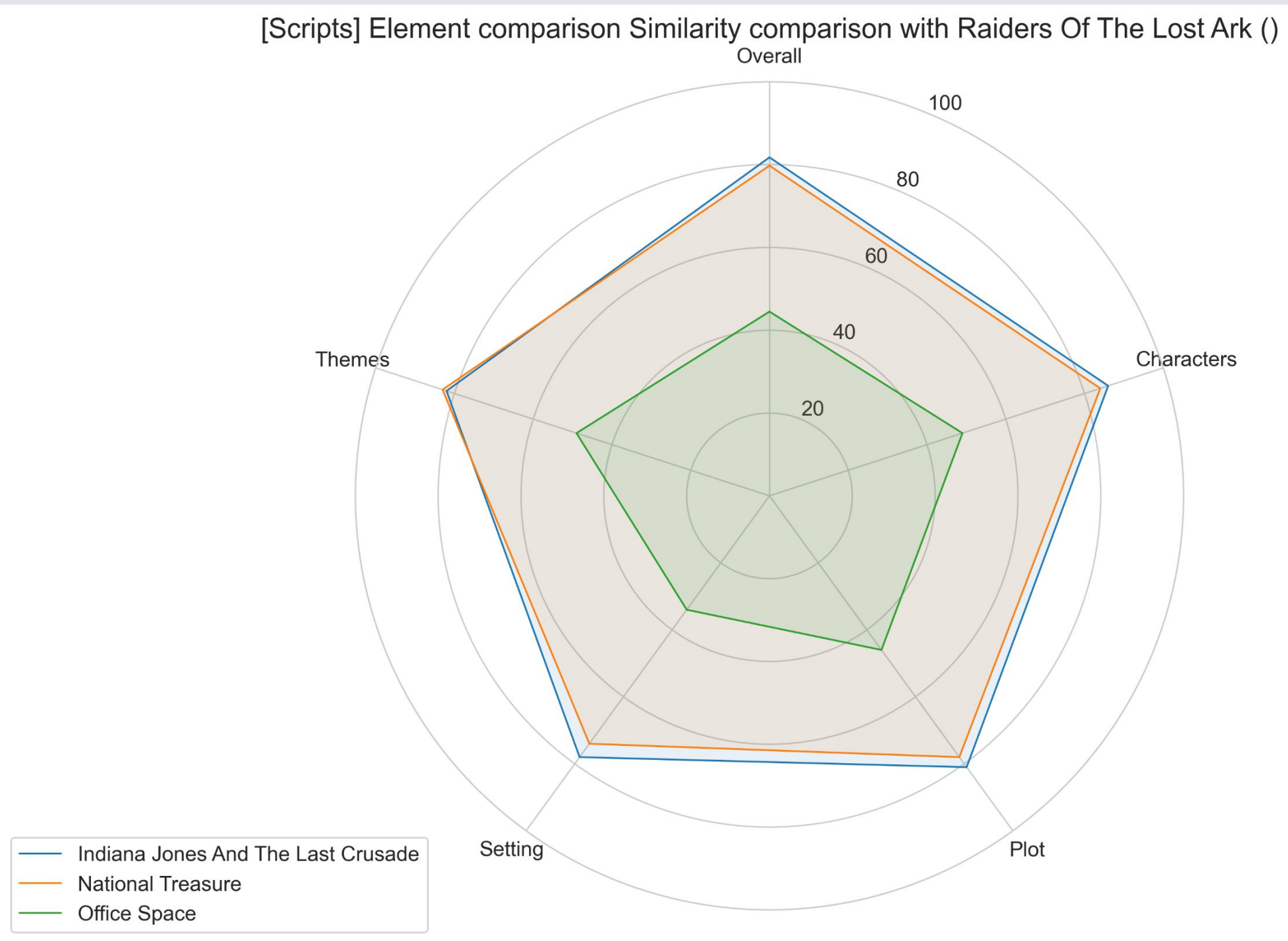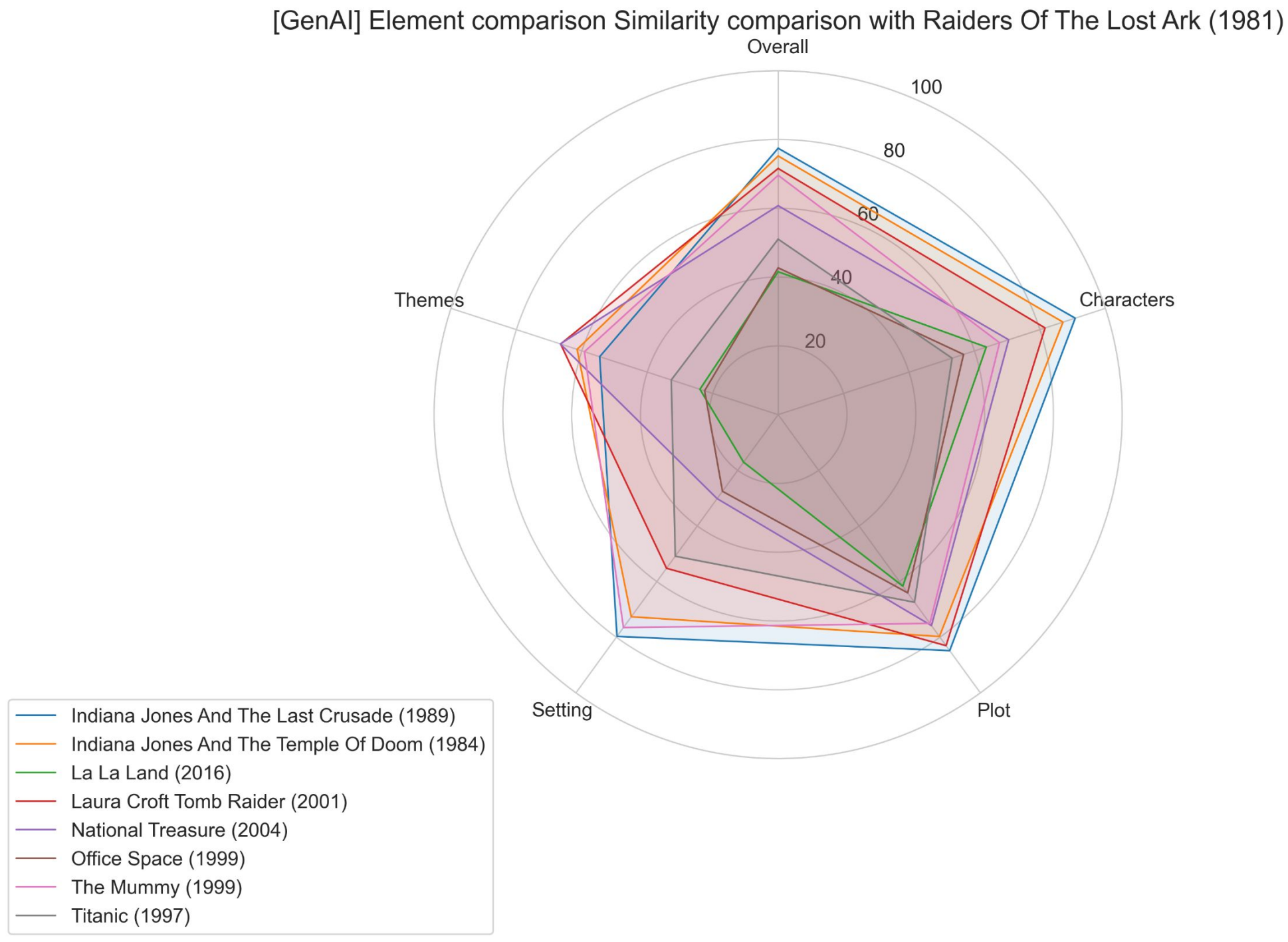

Figure 2. Narrative Elements and Sub-Features

Results

3 Methods Compared

GenAI vs Script Methods

Overall and Narrative Component Scores

Human Similarity-Title		1-Temple of Doom	2-Last Crusade	3-Tomb Raider	4-The Mummy	5-National Treasure	6-Titanic	7-Office Space	8-La La Land
Similarity Method	Narrative Element	1984 Sequel	1989 Sequel	Adventure	Adventure	Adventure	Drama-Romance	Black Comedy	Musical
Elements	Overall	70.96 (2)	73.05 (1)	66.84 (6)	69.34 (4)	68.86 (5)	69.85 (3)	62.94 (8)	63.38 (7)
	Characters	80.00	80.00	80.00	80.00	80.00	80.00	80.00	80.00
	Plot	72.50 (4)	71.62 (5)	70.00 (7)	77.06 (1)	76.62 (2)	70.88 (6)	70.00 (8)	73.53 (3)
	Setting	58.82 (3)	70.00 (1)	40.00 (tie 5-8)	60.29 (2)	40.00 (tie 5-8)	57.65 (4)	40.00 (tie 5-8)	40.00 (tie 5-8)
	Themes	72.50 (3)	70.59 (5)	77.35 (2)	60.00 (tie 6-8)	78.82 (1)	70.88 (4)	61.76 (tie 6-8)	60.00 (tie 6-8)
GenAI	Overall	75.22 (2)	77.46 (1)	71.58 (3)	69.61 (4)	60.76 (5)	51.03 (6)	42.69 (7)	41.56 (8)
	Characters	87.06 (2)	90.79 (1)	81.58 (3)	67.67 (5)	70.45 (4)	53.15 (8)	56.67 (7)	63.61 (6)
	Plot	79.73 (3)	84.82 (1)	83.03 (2)	75.03 (5)	75.82 (4)	67.42 (6)	64.06 (7)	61.64 (8)
	Setting	72.64 (3)	79.70 (1)	55.21 (4)	76.52 (2)	30.18 (6)	50.88 (5)	27.55 (7)	17.09 (8)
	Themes	61.45 (2)	54.55 (5)	66.48 (2)	59.21 (4)	66.58 (1)	32.67 (6)	22.48 (8)	23.91 (7)
Scripts	Overall		81.75 (1)			79.75 (2)		44.50 (3)	
	Characters		86.00 (1)			84.00 (2)		49.00 (3)	
	Plot		81.00 (1)			78.00 (2)		46.00 (3)	
	Setting		78.00 (1)			74.00 (2)		34.00 (3)	
	Themes		82.00 (2)			83.00 (1)		49.00 (3)	

Discussion

The benchmark has 3 groupings of films ordered by a human expert by increasing narrative distance: sequels, adventure genre and non-adventure genres. Overall AIStorySimiliarity scores largely follow this expected pattern.

The Elements method loses subtle differences and narrows overall semantic differences because it first distills narratives into concise features. This leads to the greatest disagreement (37.5%) with human experts.

In contrast, the GenAI and Script methods have only 6.25% and 17% disagreement with the human expert respectively when normalized. The Script method relies on a reference script and is more suitable for IP infringement. The GenAI method depends upon the reference script well represented in the LLM training data, so it is not suitable for comparing new or very recent narratives.

Of the 4 major narrative elements, plot is the most similar across all films likely due to the strong 'hero's journey' narrative in major Hollywood films. In contrast, Themes is the most dissimilar possibly since it is the most abstract,

Discussion (cont.)

(cont.) subjective, and artistically unconstrained of the four narrative elements measured.

A human excerpt chose the benchmark film dataset to have clear similarity clustering by sequels, adventure genre, and non-adventure genre. This is clearly reflected in AIStorySimilarity's overall narrative similarity metric. The GenAI radar plot clearly shows three overall narrative distance along the noon spoke clusters into sequels, adventure genres, and non-adventure genres.

References

1. Snyder, B. (2005). Save the Cat! The Last Book on Screenwriting You'll Ever Need. United States: Michael Wiese Productions.
2. https://semeval.github.io/SemEval2022/tasks.html
3. Ippolito, Daphne, et al. "Creative writing with an ai-powered writing assistant: Perspectives from professional writers." *arXiv preprint arXiv:2211.05030* (2022).
4. Wang, Jiaan, et al. "Is chatgpt a good NLG evaluator? a preliminary study." arXiv preprint arXiv:2303.04048 (2023).
5. https://github.com/ jon-chun/AIStorySimiliarity