

Text Analysis Report: marvel_cinematic_universe_chatgpt

1. Basic Statistics

Metric	Value
Word Count	382
Sentence Count	25
Character Count	2613
Unique Word Count	208
Avg Word Length (chars)	5.62
Avg Sentence Length (words)	15.28
Avg Sentence Length (chars)	104.52
Sentence Length Std Dev (words)	8.74
Lexical Diversity (TTR)	0.5445
MATTR (Window=50)	0.8485

2. Entropy & Complexity

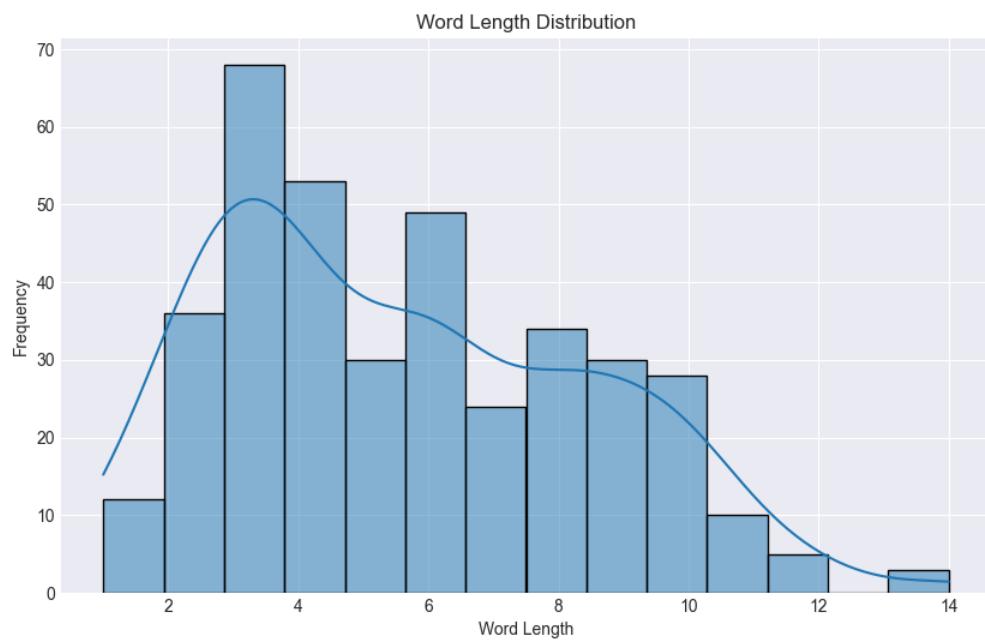
Metric	Value
Shannon Entropy (Words)	7.1437 bits
Shannon Entropy (Chars)	4.7231 bits
Function Word Ratio	0.2827

3. Readability Scores

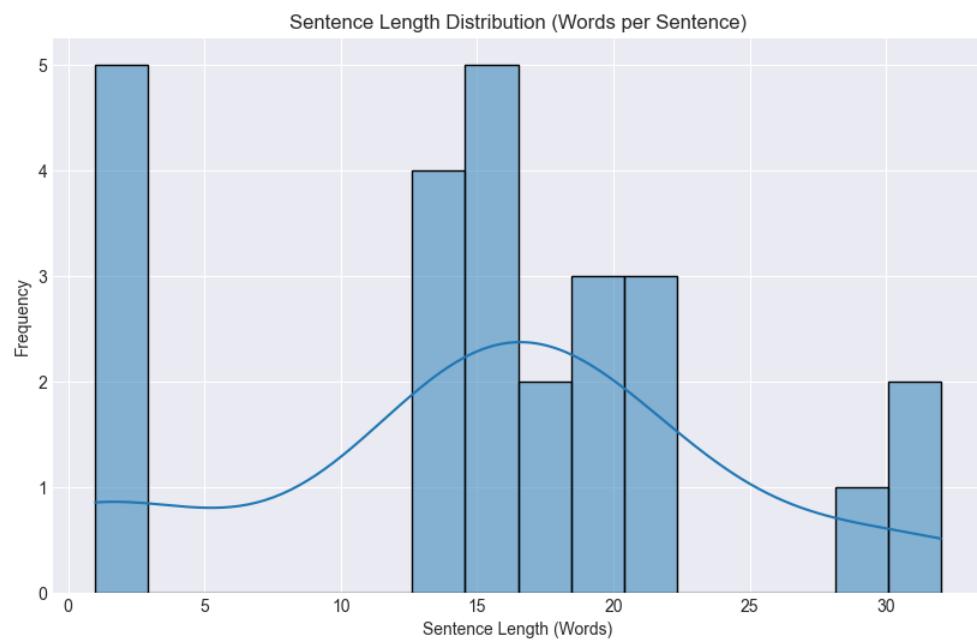
Index	Score
Flesch Reading Ease	27.52
Flesch-Kincaid Grade Level	14.00
Gunning Fog Index	14.00
SMOG Index	15.50
Coleman-Liau Index	16.82
Automated Readability Index	16.70

4. Visualizations

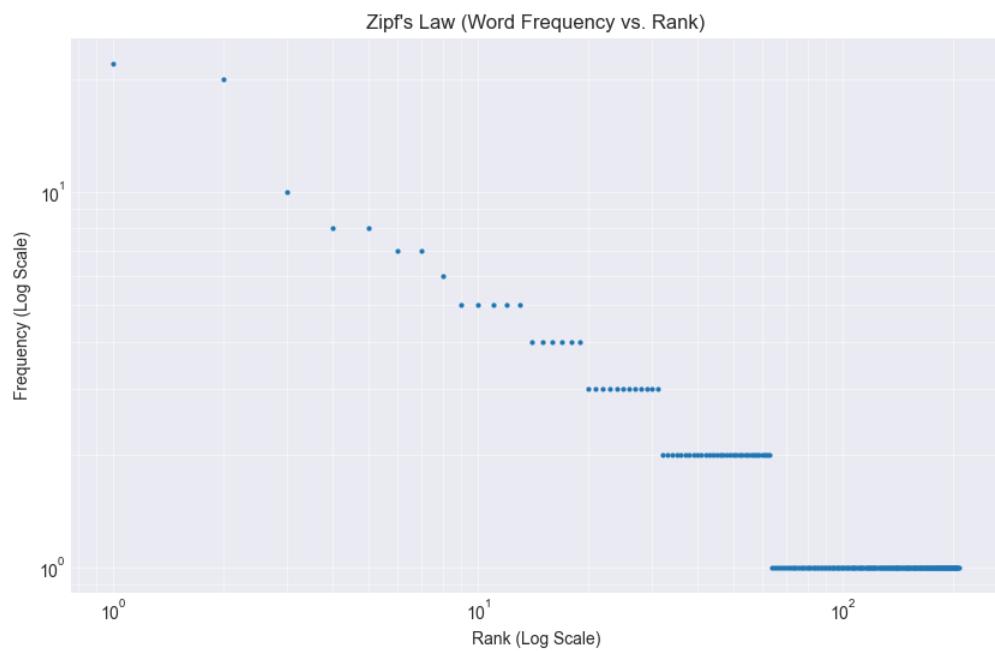
Word Length Distribution



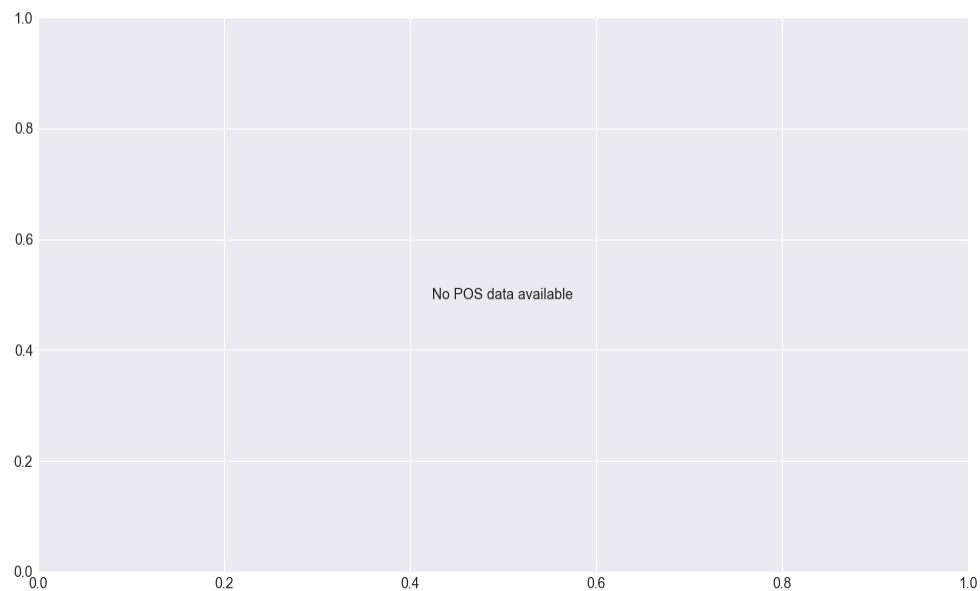
Sentence Length Distribution



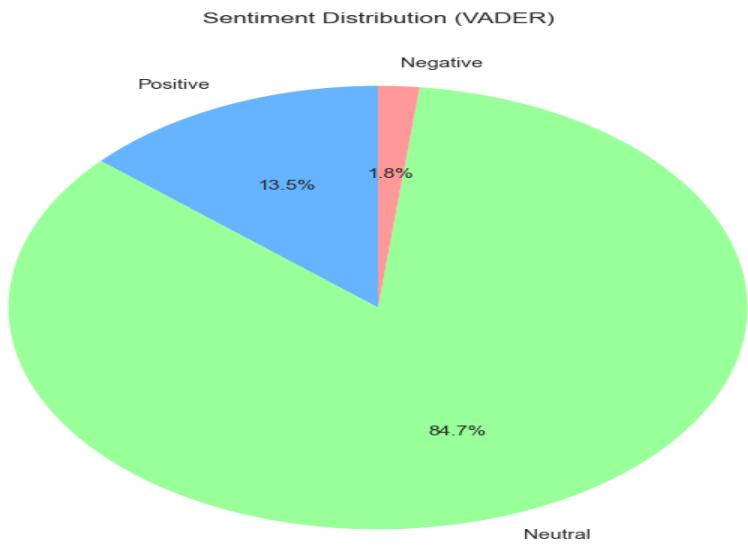
Zipf's Law Analysis (Word Frequency vs Rank)



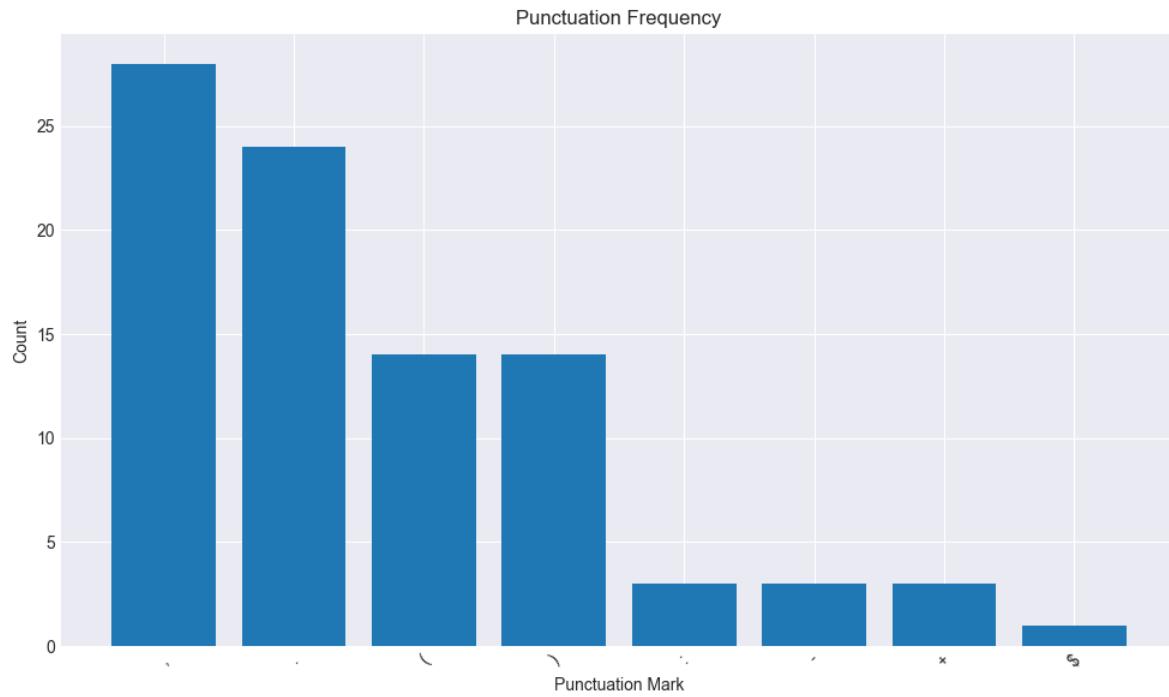
Part-of-Speech Distribution (Top 25)



Sentiment Distribution (VADER)



Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (22), and (20), series (10), of (8), films (8), marvel (7), with (7), avengers (6), universe (5), mcu (5), franchise (5), to (5), phase (5), is (4), a (4)

Most Common Punctuation:

',' (28), '.' (24), '(' (14), ')' (14), '-' (3), ':' (3), '+' (3), '\$' (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'mcu is' (3), 'films and' (3), 'the mcu' (3), 'marvel cinematic' (2), 'cinematic universe' (2), 'universe mcu' (2), 'superhero films' (2), 'television series' (2), 'marvel studios' (2), 'marvel comics' (2)

2-gram Repetition Rate: 0.0814

Common 3-grams (Top 10):

'marvel cinematic universe' (2), 'cinematic universe mcu' (2), 'is known for' (2), 'multiple films and' (2), 'films and series' (2), 'the mcu is' (2), 'expanded the universe' (2), 'universe mcu the' (1), 'mcu the marvel' (1), 'the marvel cinematic' (1)

3-gram Repetition Rate: 0.0184

Common 4-grams (Top 10):

'marvel cinematic universe mcu' (2), 'multiple films and series' (2), 'cinematic universe mcu the' (1), 'universe mcu the marvel' (1), 'mcu the marvel cinematic' (1), 'the marvel cinematic universe' (1), 'cinematic universe mcu is' (1), 'universe mcu is a' (1), 'mcu is a media' (1), 'is a media franchise' (1)

4-gram Repetition Rate: 0.0053

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'and' (25), 'the' (23), 'ed_' (23), 'he_' (22), 'nd_' (22), '_th' (22), 'ing' (22), '_an' (21), 'es_' (21), 'ng_' (20)

Common Character 4-grams (Top 10):

'the_' (22), '_and' (20), 'and_' (20), 'ing_' (20), '_the' (18), 'ries' (11), '_int' (11), '_fil' (11), 'film' (11), 'vers' (10)

Common Character 5-grams (Top 10):

'_and_' (20), '_the_' (18), '_film' (11), '_seri' (10), 'serie' (10), 'eries' (10), 'ivers' (9), 'verse' (9), 'ries_' (9), 's_and' (9)