

Text Analysis Report: crispr_gene_editing_chatgpt

1. Basic Statistics

Metric	Value
Word Count	204
Sentence Count	10
Character Count	1368
Unique Word Count	138
Avg Word Length (chars)	5.52
Avg Sentence Length (words)	20.40
Avg Sentence Length (chars)	136.80
Sentence Length Std Dev (words)	6.31
Lexical Diversity (TTR)	0.6765
MATTR (Window=50)	0.8386

2. Entropy & Complexity

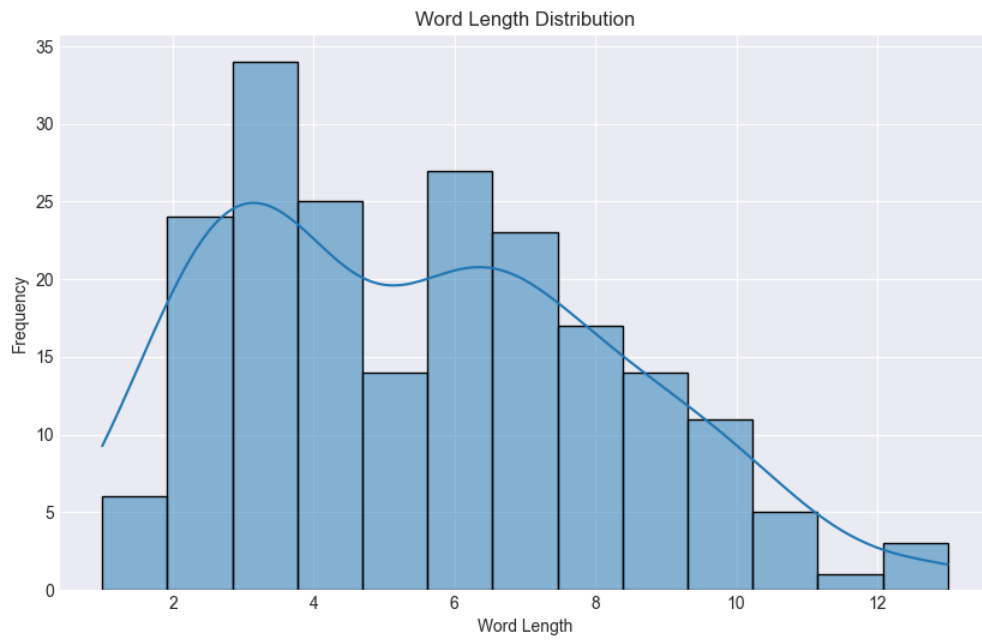
Metric	Value
Shannon Entropy (Words)	6.7472 bits
Shannon Entropy (Chars)	4.6751 bits
Function Word Ratio	0.2794

3. Readability Scores

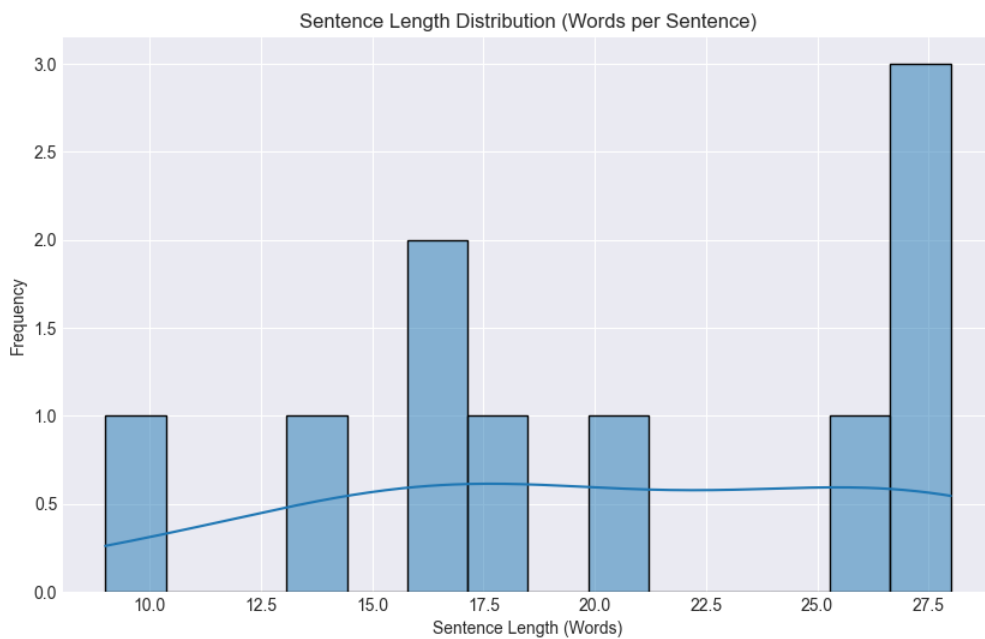
Index	Score
Flesch Reading Ease	17.54
Flesch-Kincaid Grade Level	15.70
Gunning Fog Index	16.81
SMOG Index	17.40
Coleman-Liau Index	15.72
Automated Readability Index	16.30

4. Visualizations

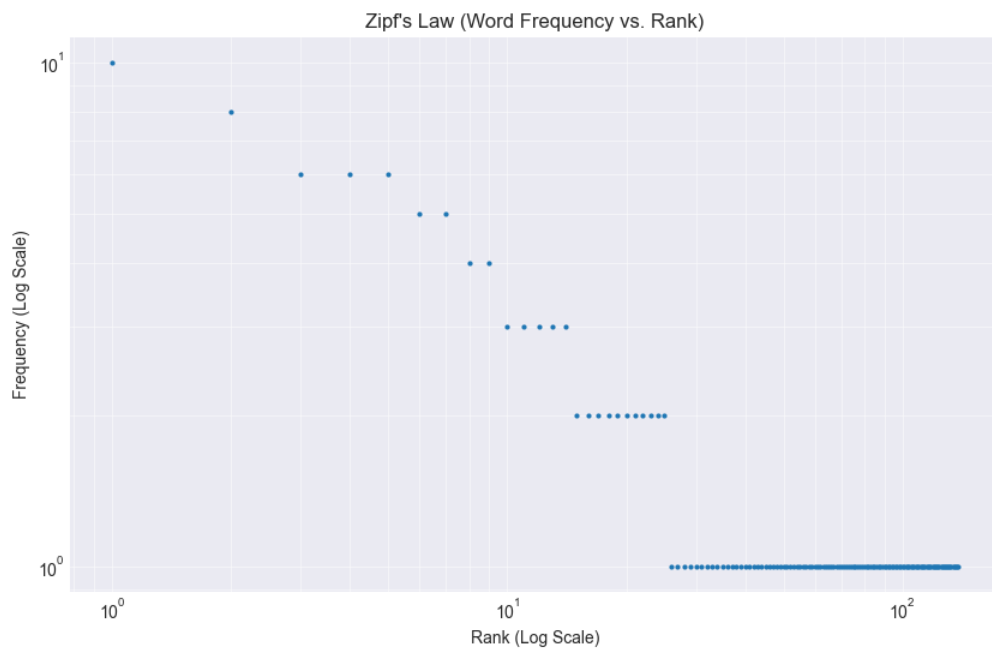
Word Length Distribution



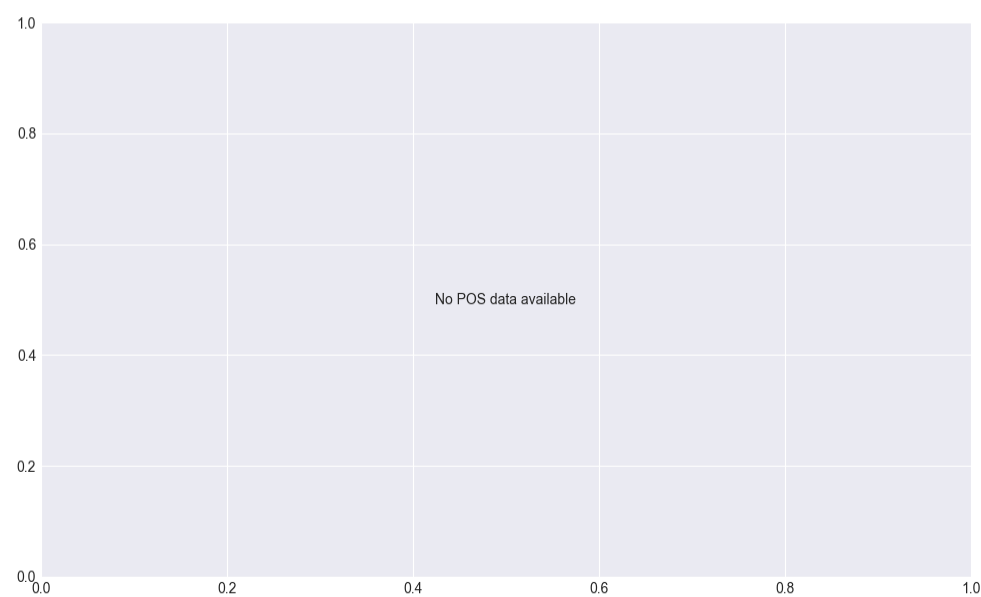
Sentence Length Distribution



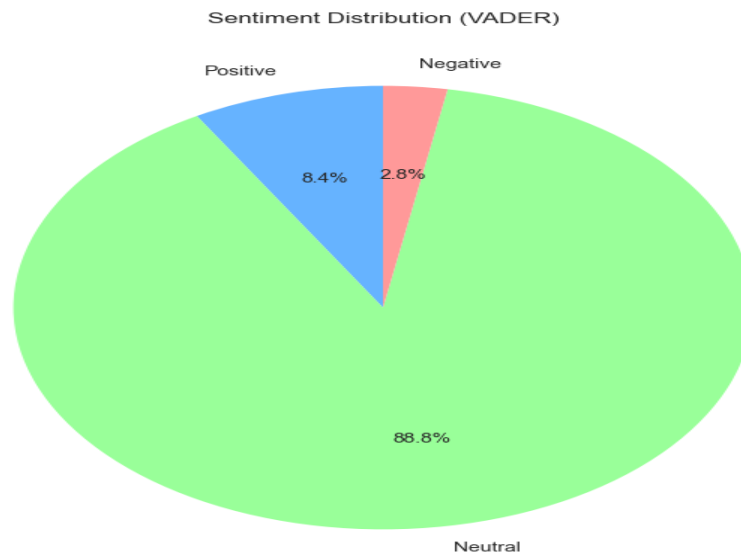
Zipf's Law Analysis (Word Frequency vs Rank)



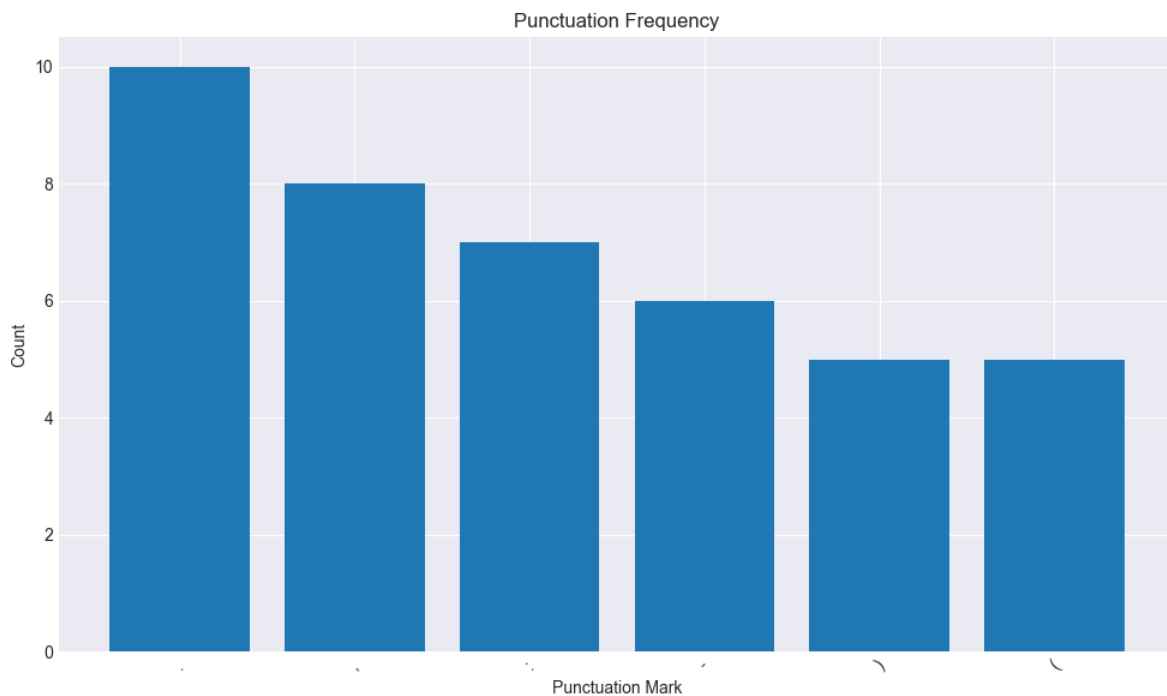
Part-of-Speech Distribution (Top 25)



Sentiment Distribution (VADER)



Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (10), crispr (8), dna (6), a (6), in (6), gene (5), to (5), editing (4), cas9 (4), of (3), repair (3), that (3), genetic (3), and (3), mechanism (2)

Most Common Punctuation:

' (10), ',' (8), ':' (7), '-' (6), '(' (5), ')' (5)

6. N-gram Analysis

Common 2-grams (Top 10):

'gene editing' (4), 'crispr gene' (2), 'cas9 to' (2), 'dna at' (2), 'one of' (2), 'to precisely' (2), 'mechanism of' (1), 'of action' (1), 'action crispr' (1), 'editing relies' (1)
2-gram Repetition Rate: 0.0394

Common 3-grams (Top 10):

'crispr gene editing' (2), 'mechanism of action' (1), 'of action crispr' (1), 'action crispr gene' (1), 'gene editing relies' (1), 'editing relies on' (1), 'relies on the' (1), 'on the cas' (1), 'the cas crispr' (1), 'cas crispr associated' (1)
3-gram Repetition Rate: 0.0050

Common 4-grams (Top 10):

'mechanism of action crispr' (1), 'of action crispr gene' (1), 'action crispr gene editing' (1), 'crispr gene editing relies' (1), 'gene editing relies on' (1), 'editing relies on the' (1), 'relies on the cas' (1), 'on the cas crispr' (1), 'the cas crispr associated' (1), 'cas crispr associated proteins' (1)
4-gram Repetition Rate: 0.0000

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'_th' (15), 'the' (12), 'ed_' (12), '_in' (11), 'ing' (10), 'he_' (10), 'na_' (10), 'ng_' (9), '_re' (9), 'cri' (8)

Common Character 4-grams (Top 10):

'_the' (12), 'the_' (10), 'ing_' (9), 'cris' (8), 'risp' (8), 'ispr' (8), '_gen' (8), 'gene' (8), 'dna_' (7), 'tion' (6)

Common Character 5-grams (Top 10):

'_the_' (10), 'crisp' (8), 'rispr' (8), '_gene' (8), 'ispr_' (6), 'anism' (4), ' cris' (4), 'editi' (4), 'ditin' (4), 'iting' (4)