

Text Analysis Report: crispr_gene_editing_bard

1. Basic Statistics

Metric	Value
Word Count	281
Sentence Count	14
Character Count	1713
Unique Word Count	140
Avg Word Length (chars)	4.88
Avg Sentence Length (words)	20.07
Avg Sentence Length (chars)	122.36
Sentence Length Std Dev (words)	8.06
Lexical Diversity (TTR)	0.4982
MATTR (Window=50)	0.7860

2. Entropy & Complexity

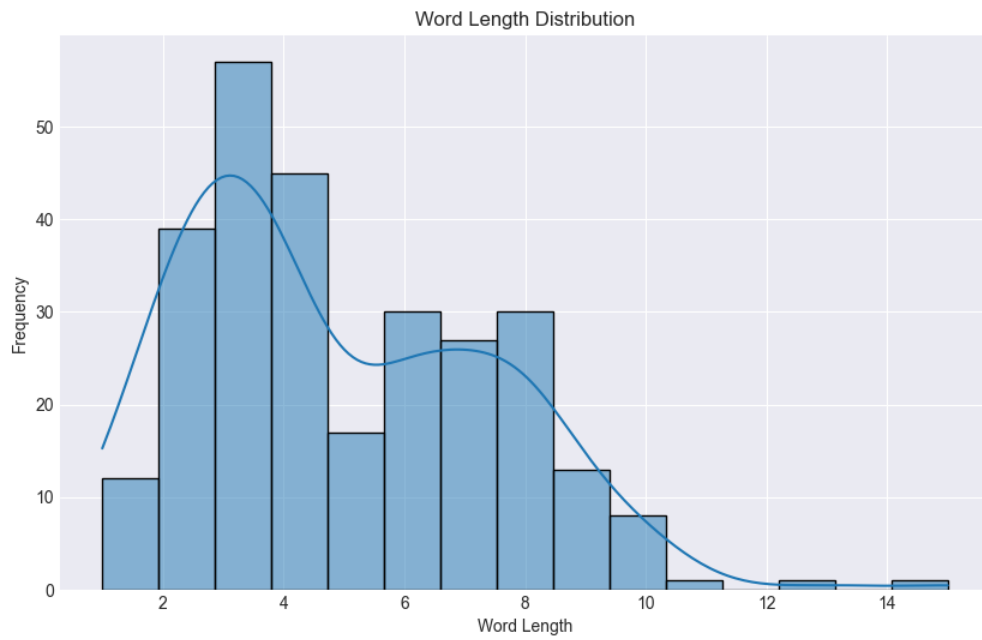
Metric	Value
Shannon Entropy (Words)	6.5664 bits
Shannon Entropy (Chars)	4.6630 bits
Function Word Ratio	0.3345

3. Readability Scores

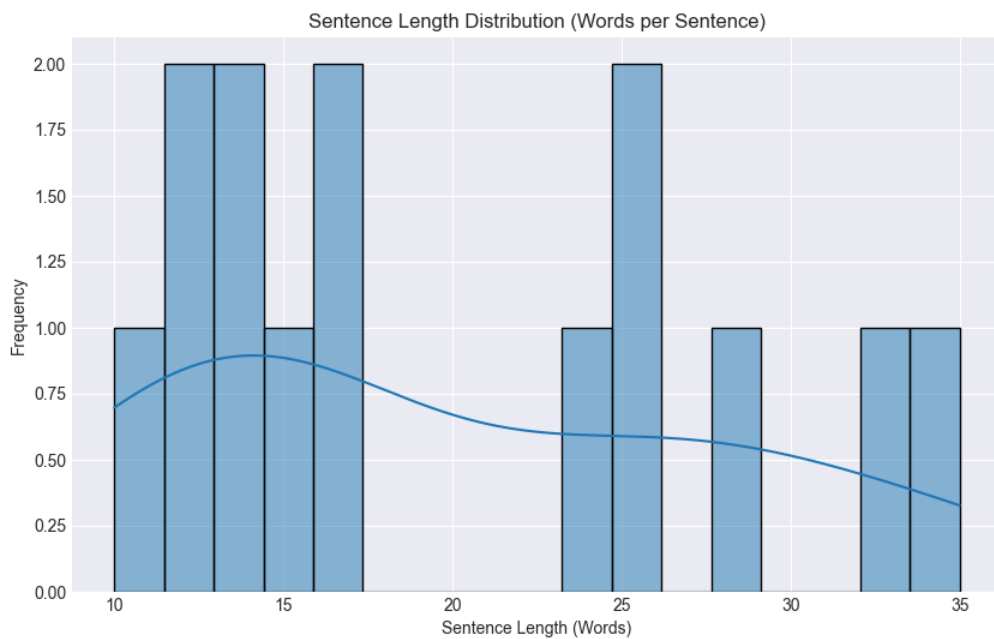
Index	Score
Flesch Reading Ease	34.86
Flesch-Kincaid Grade Level	13.20
Gunning Fog Index	13.66
SMOG Index	15.00
Coleman-Liau Index	12.07
Automated Readability Index	13.30

4. Visualizations

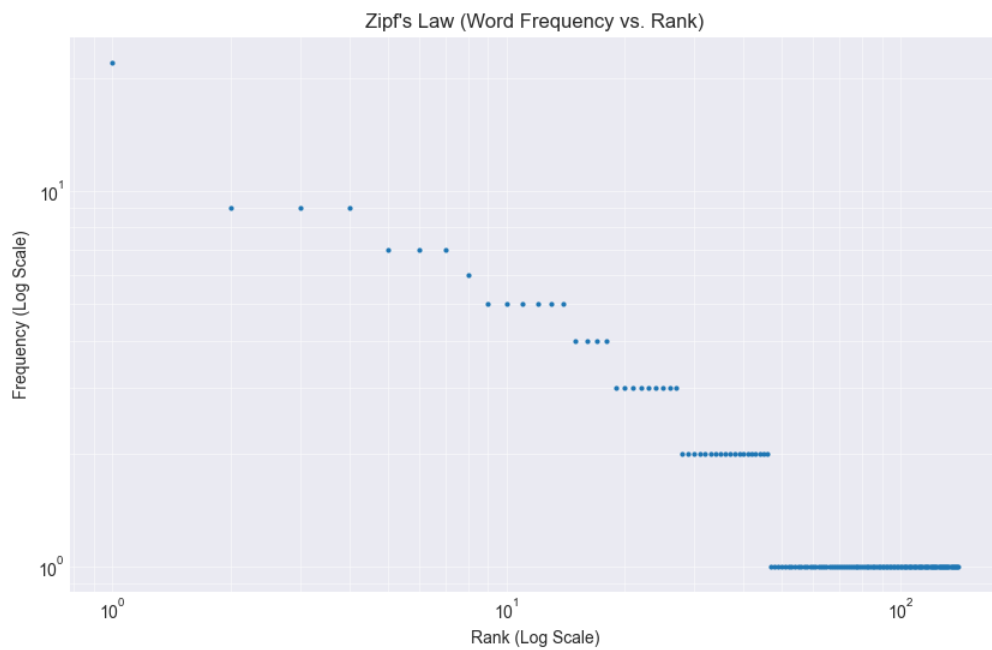
Word Length Distribution

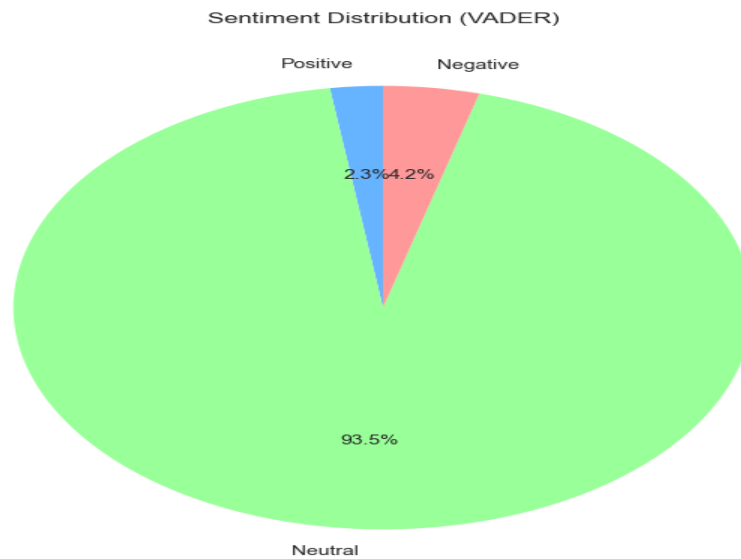


Sentence Length Distribution

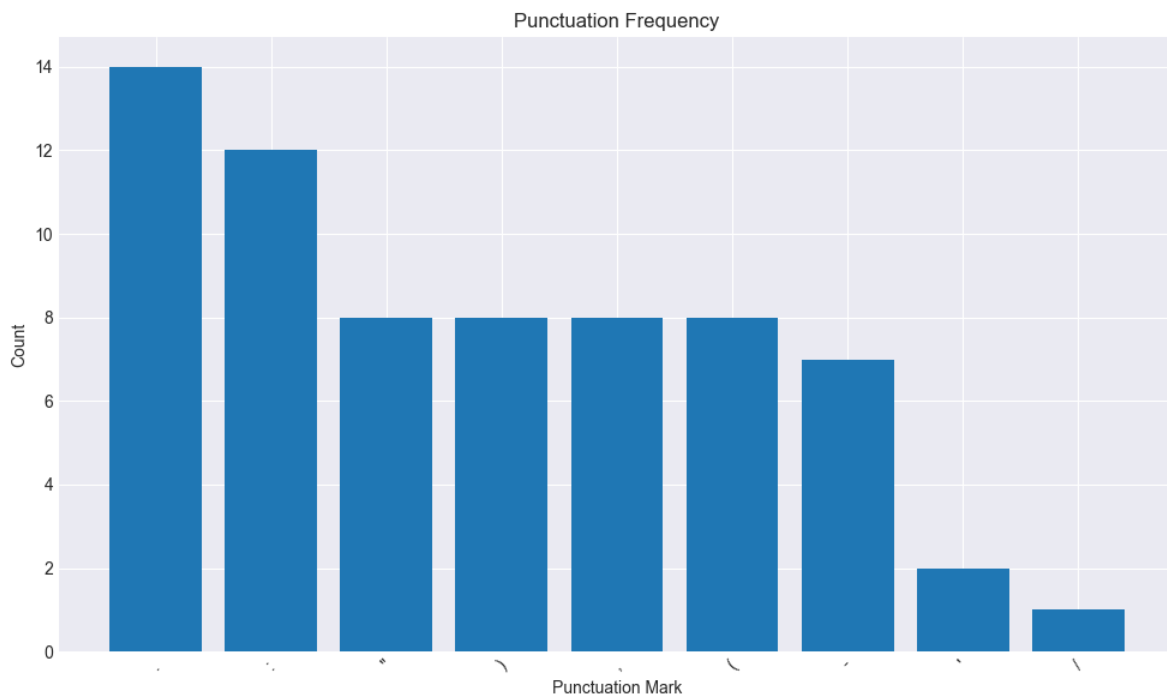


Zipf's Law Analysis (Word Frequency vs Rank)





Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (22), a (9), cas9 (9), dna (9), gene (7), grna (7), to (7), crispr (6), is (5), of (5), be (5), target (5), sequence (5), repair (5), editing (4)

Most Common Punctuation:

' (14), ':' (12), '(' (8), ')' (8), ',' (8), '"' (8), '-' (7), '=' (2), '/' (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'gene editing' (4), 'dna sequence' (4), 'the target' (4), 'crispr cas9' (3), 'the cas9' (3), 'to the' (3), 'target dna' (3), 'crispr gene' (2), 'of the' (2), 'guide rna' (2)
2-gram Repetition Rate: 0.1250

Common 3-grams (Top 10):

'crispr gene editing' (2), 'guide rna grna' (2), 'target dna sequence' (2), 'to the target' (2), 'the target dna' (2), 'allowing for precise' (2), 'specific dna sequence' (2), 'the cas9 protein' (2), 'gene editing crispr' (1), 'editing crispr gene' (1)
3-gram Repetition Rate: 0.0287

Common 4-grams (Top 10):

'to the target dna' (2), 'crispr gene editing crispr' (1), 'gene editing crispr gene' (1), 'editing crispr gene editing' (1), 'crispr gene editing is' (1), 'gene editing is a' (1), 'editing is a genetic' (1), 'is a genetic engineering' (1), 'a genetic engineering technique' (1), 'genetic engineering technique in' (1)
4-gram Repetition Rate: 0.0036

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'_th' (26), 'the' (22), 'he_' (22), 'ing' (19), 'na_' (18), 'ng_' (17), 'ed_' (12), '_ge' (11), 'gen' (11), 'tin' (10)

Common Character 4-grams (Top 10):

'the_' (22), '_the' (20), 'ing_' (17), '_gen' (11), 'ting' (10), 'rna_' (10), 'gene' (9), 'cas9' (9), 'as9_' (9), '_dna' (9)

Common Character 5-grams (Top 10):

'_the_' (20), '_gene' (9), 'cas9_' (9), 'ting_' (8), 'the_c' (8), '_dna_' (8), 'gene_' (7), 'grna_' (7), 'targe' (7), 'arget' (7)