

# Text Analysis Report: crispr\_gene\_editing\_deepseek

## 1. Basic Statistics

Metric	Value
Word Count	309
Sentence Count	15
Character Count	2005
Unique Word Count	174
Avg Word Length (chars)	5.29
Avg Sentence Length (words)	20.60
Avg Sentence Length (chars)	133.67
Sentence Length Std Dev (words)	5.14
Lexical Diversity (TTR)	0.5631
MATTR (Window=50)	0.8228

## 2. Entropy & Complexity

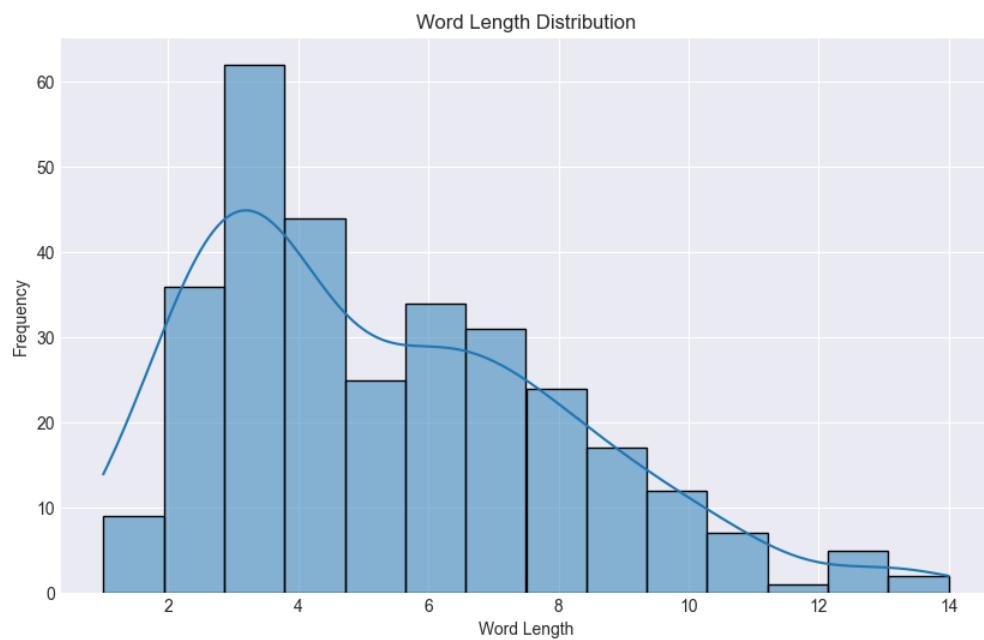
Metric	Value
Shannon Entropy (Words)	6.9360 bits
Shannon Entropy (Chars)	4.6623 bits
Function Word Ratio	0.3139

## 3. Readability Scores

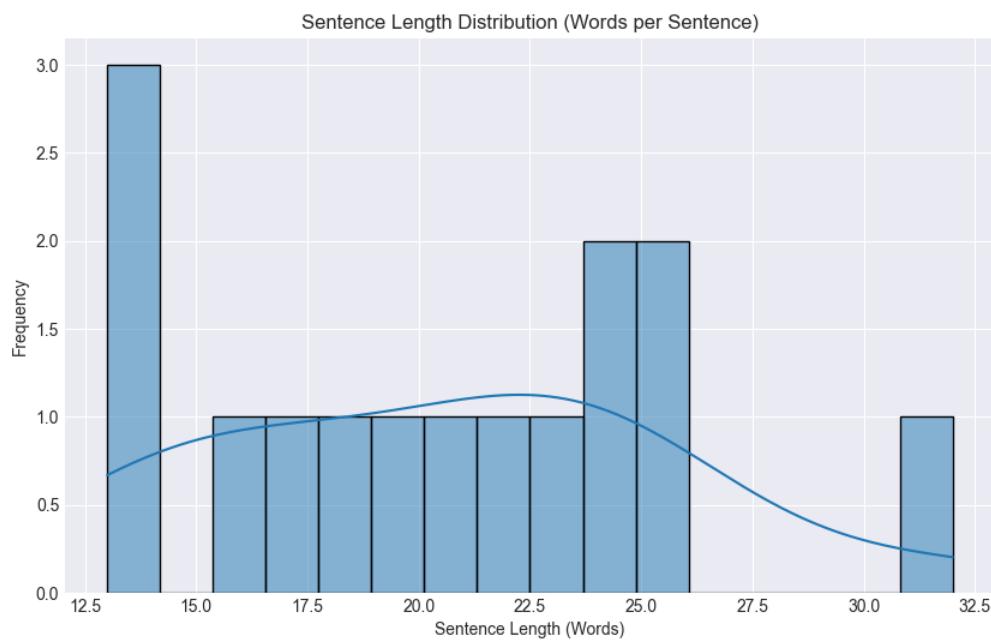
Index	Score
Flesch Reading Ease	25.69
Flesch-Kincaid Grade Level	14.70
Gunning Fog Index	15.48
SMOG Index	16.80
Coleman-Liau Index	14.27
Automated Readability Index	15.10

## 4. Visualizations

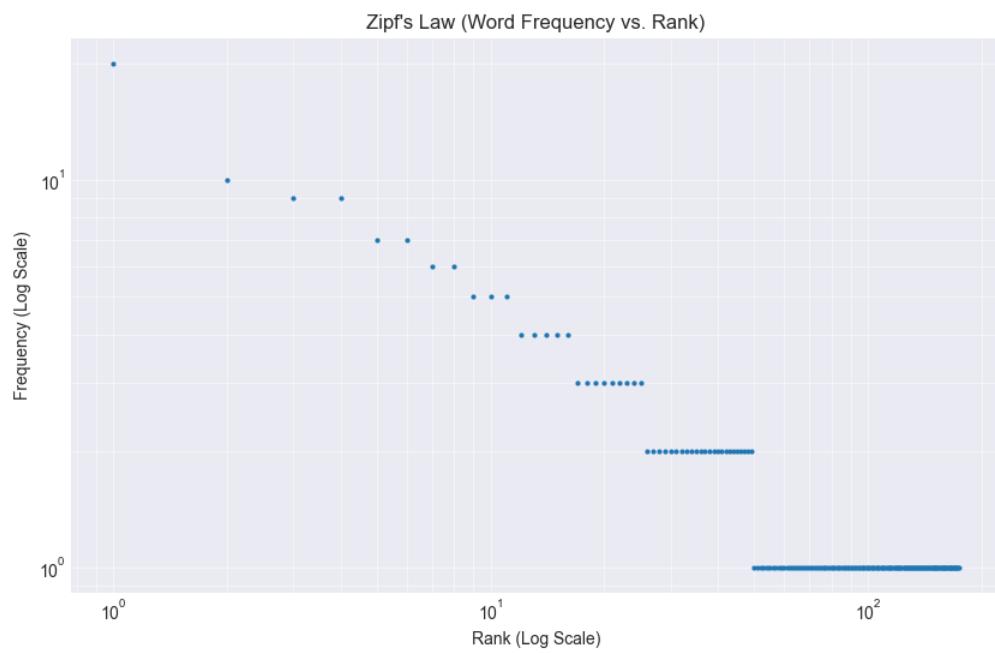
### Word Length Distribution



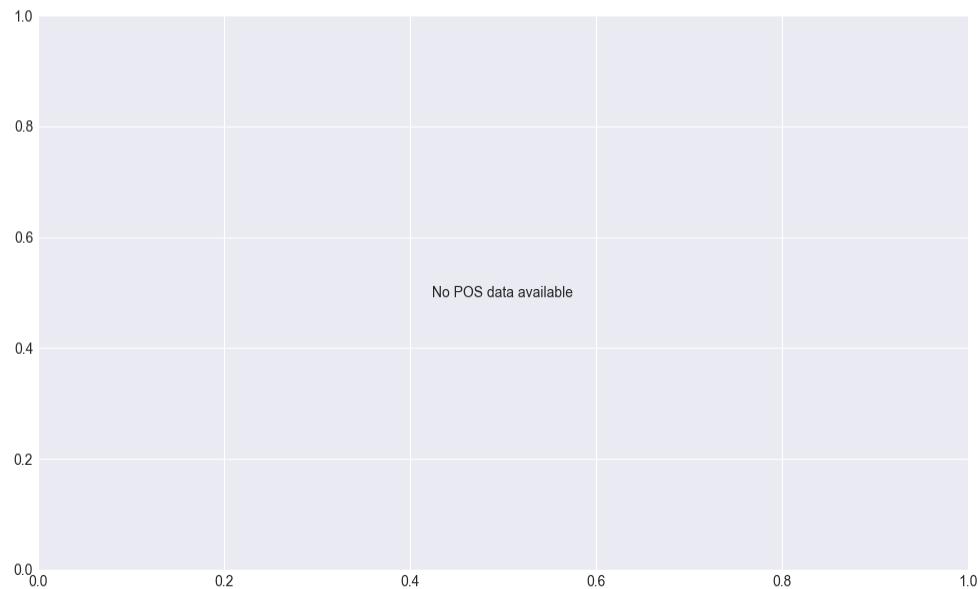
## Sentence Length Distribution



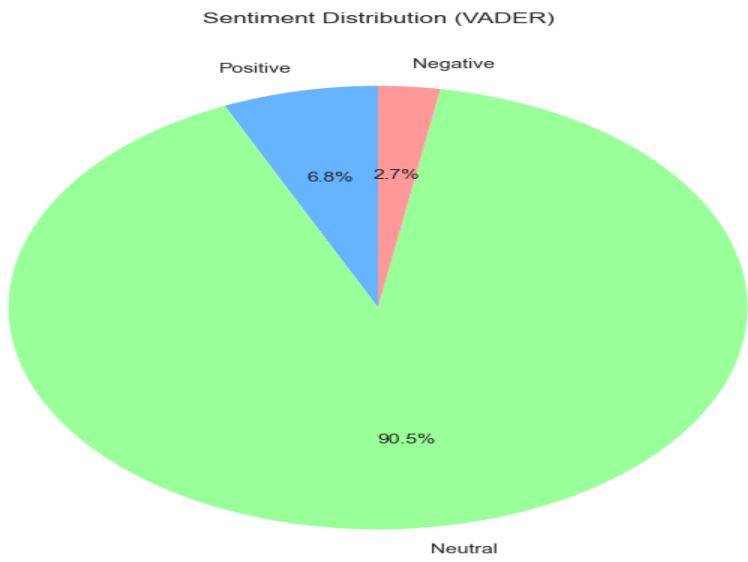
## Zipf's Law Analysis (Word Frequency vs Rank)



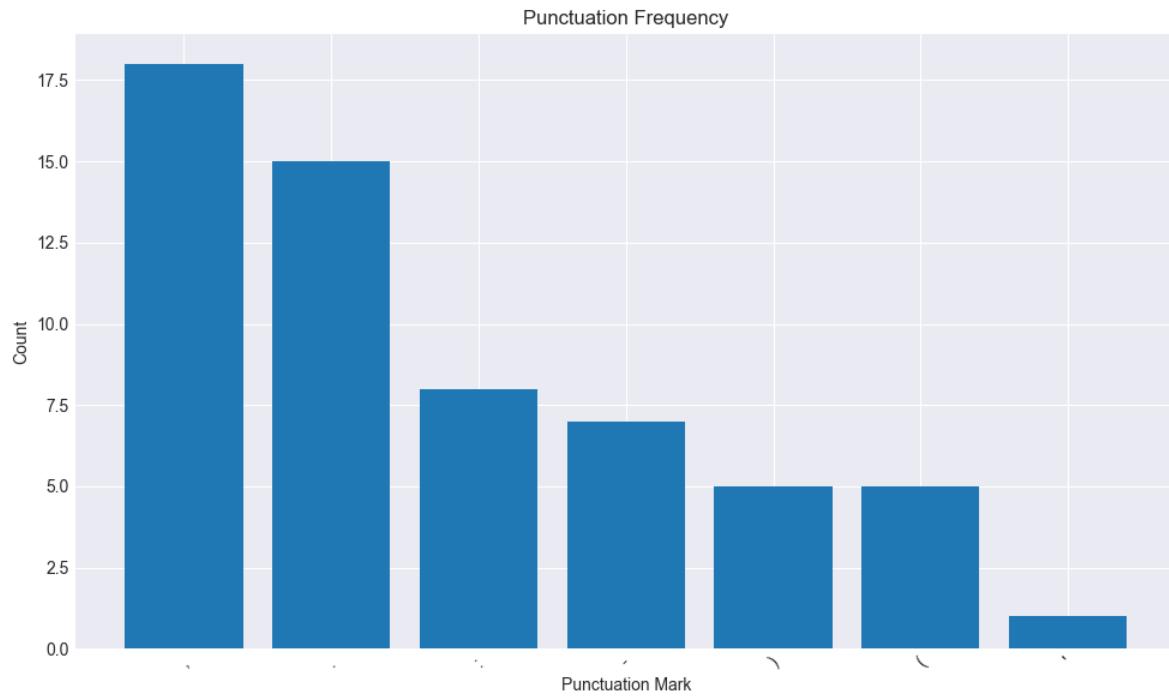
## Part-of-Speech Distribution (Top 25)



## Sentiment Distribution (VADER)



## Punctuation Frequency



## 5. Frequency Analysis

### Most Common Words (Top 15, with stopwords):

the (20), dna (10), crispr (9), to (9), a (7), cas9 (7), in (6), and (6), gene (5), of (5), grna (5), editing (4), for (4), genetic (4), rna (4)

## **Most Common Punctuation:**

',' (18), '.' (15), ':' (8), '-' (7), '(' (5), ')' (5), '"' (1)

## **6. N-gram Analysis**

### **Common 2-grams (Top 10):**

'gene editing' (4), 'crispr gene' (3), 'the dna' (3), 'cas9 enzyme' (3), 'the cas9' (3), 'the target' (3), 'is a' (2), 'the crispr' (2), 'dna at' (2), 'guide rna' (2)

2-gram Repetition Rate: 0.0974

### **Common 3-grams (Top 10):**

'crispr gene editing' (3), 'the cas9 enzyme' (2), 'the target dna' (2), 'target dna sequence' (2), 'cas9 grna complex' (2), 'binds to the' (2), 'gene editing crispr' (1), 'editing crispr gene' (1), 'gene editing is' (1), 'editing is a' (1)

3-gram Repetition Rate: 0.0228

### **Common 4-grams (Top 10):**

'the target dna sequence' (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 revolutionary' (1), 'is a revolutionary biotechnology' (1), 'a revolutionary biotechnology that' (1), 'revolutionary biotechnology that allows' (1)

4-gram Repetition Rate: 0.0033

## **Character N-gram Analysis:**

### **Common Character 3-grams (Top 10):**

'\_th' (24), 'the' (23), 'he\_' (20), 'na\_' (20), 'ing' (16), 'ed\_' (16), 'ng\_' (15), '\_in' (14), 'e\_c' (13), '\_ge' (12)

### **Common Character 4-grams (Top 10):**

'\_the' (22), 'the\_' (20), 'ing\_' (15), '\_gen' (12), 'dna\_' (11), 'cris' (9), 'risp' (9), 'ispr' (9), 'gene' (9), '\_to\_' (9)

### **Common Character 5-grams (Top 10):**

'\_the\_' (19), 'crisp' (9), 'rispr' (9), '\_gene' (9), 'the\_c' (8), 'ting\_' (7), '\_dna\_' (7), 'ispr\_' (6), '\_cris' (6), '\_and\_' (6)