AI VS HUMAN CONTENT DETECTION 1000+ RECORD IN 2025

AN ANALYSIS CONDUCTED ON THIS KAGGLE DATASET USING PYTHON (PANDAS) AND POWER BI

Sabastian Ihekweme
sabastianfabian001@gmail.com |
https://sabastian-ihekweme.vercel.app/

September 2025

TABLE OF CONTENT

1. INTRODUCTION TO THE PROJECT

This project is my first comprehensive analysis using Python and Pandas, and I used the data wrangling, exploratory analysis, and summary table building approaches that I had previously acquired. I worked on the "AI vs Human Content Detection" Kaggle dataset (1,000+ records, 2025), which is an interesting and relevant topic since finding linguistic variations between AI generated and human written text will help refine content detection models and improve overall AI detection performance.

In addition to extensive data cleaning, I created several summary tables in Python to capture crucial statistics and trends before developing interactive Power BI dashboards. This combination of structured Python analysis and visual reporting demonstrates how Pandas and Power BI can extract significant insights from raw text data.

2. OVERVIEW OF THE DATA

The research is based on the "AI vs Human Content Detection" Kaggle dataset (2025) which includes over 1,000 text samples with precise linguistic and stylistic attributes. Each record contains both raw text and 16 constructed attributes that measure readability, sentence organization, word richness, and grammatical accuracy.

This dataset is particularly fascinating since it combines natural language processing and AI ethics. As generative AI technologies grow more common, separating human written text from machine-generated content becomes critical for academic integrity, media verification, and automated moderation systems. Analysts can improve AI content identification models and reduce false positives or negatives by studying criteria such as lexical variety, readability scores, and predictability.

DATA DICTIONARY

Column Name	**Data Type**	**Description**	**Example Values**
text_content	String	Raw text sample to be analyzed	“Score each cause. Quality throughout...”
content_type	String	Category/genre of the content	academic_paper, essay, creative_writing
word_count	Integer	Total number of words	288, 253
character_count	Integer	Total characters including spaces	1927, 1719
sentence_count	Integer	Number of senten ces	54, 45
lexical_diversity	Float	Ratio of unique word to total words	0.95, 0.97
avg_sentence_length	Float	Average words per sentence	5.33, 5.62
avg_word_length	Float	Average characters per	5.69, 5.8

		word	
punctuation_ratio	Float	Ratio of punctuation marks to total characters	0.028, 0.026
flesch_reading_ease	Float	Readability score: higher = easier	53.1, 50.3
gunning_fog_index	Float	Gunning Fog grade level required to understand text	7.4, 8.1
grammar_error	Integer	Number of detected grammatical errors	1, 6
passive_voice_ratio	Float	Ratio of passive sentences (0-1)	0.10, 0.20
predictability_score	Float	Higher values indicate more predictable sentence	105.8, 96.9
burstiness	Float	Variation in sentence length	0.55, 0.62
sentiment_score	Float	Sentiment polarity (-1 to +1; negative to positive)	0.20, -0.23
label	Integer	Target variable: 1 = AI-Generated, 0 = Human-Written	1, 0

3. DATA CLEANING & PREPARATION

To guarantee that the dataset was accurate and ready for analysis, I used a structured data preparation workflow. The following is the narrative that you can adapt:

1. Initial Inspection

In order to verify the dataset's shape and find any obvious irregularities, I started by looking at its structure and summary statistics. I verified the amount of rows/columns, data types, and null counts, df.shape and df.info() were utilized.

```
import pandas as pd

#Data Cleaning

df = pd.read_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/dataframe.csv')

# Dataset dimension and basics info
df.shape
df.info()
```

```
✓ [23] 37ms
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1367 entries, 0 to 1366
Data columns (total 17 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   text_content         1367 non-null   object
 1   content_type         1367 non-null   object
 2   word_count           1367 non-null   int64
 3   character_count      1367 non-null   int64
 4   sentence_count       1367 non-null   int64
 5   lexical_diversity    1367 non-null   float64
 6   avg_sentence_length  1367 non-null   float64
 7   avg_word_length      1367 non-null   float64
```

2. Validation of Data Type and Column

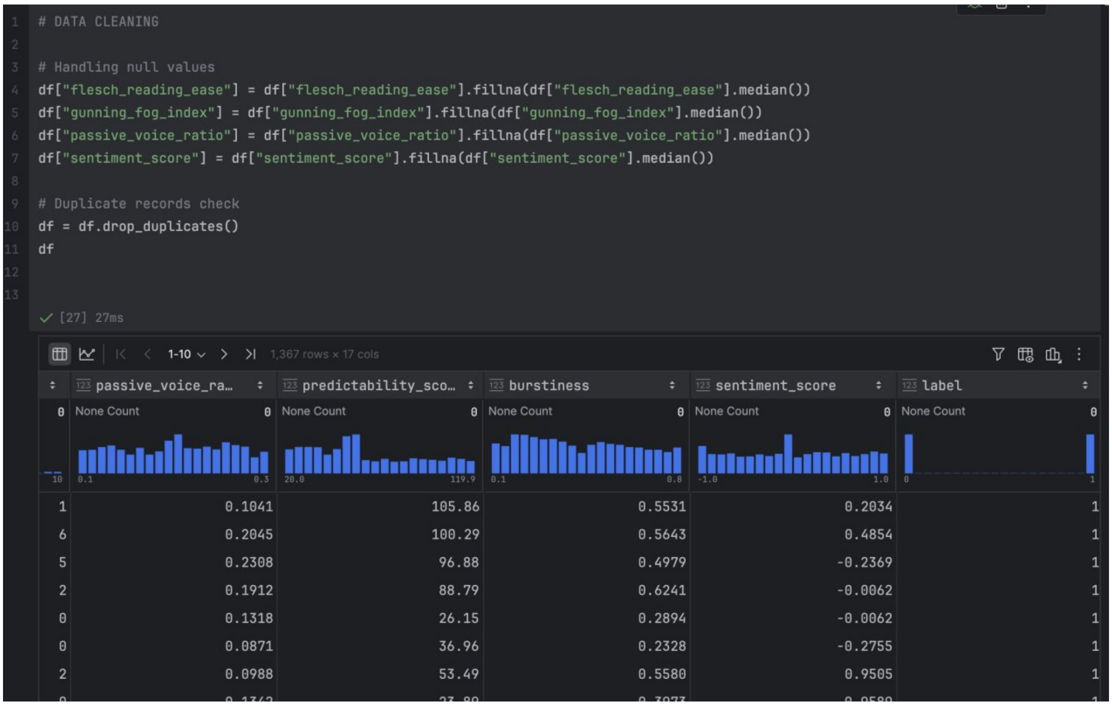
To make sure the data type matched the desired format (e.g., integers for counts, floating for ratios), each column was examined.

3. Handling Missing Values

There were missing values in a number of columns (such as sentiment_score, which occasionally had gaps).

4. Removing Duplicates

I checked for, and dropped duplicate rows to avoid bias.



```
# DATA CLEANING

# Handling null values
df["flesch_reading_ease"] = df["flesch_reading_ease"].fillna(df["flesch_reading_ease"].median())
df["gunning_fog_index"] = df["gunning_fog_index"].fillna(df["gunning_fog_index"].median())
df["passive_voice_ratio"] = df["passive_voice_ratio"].fillna(df["passive_voice_ratio"].median())
df["sentiment_score"] = df["sentiment_score"].fillna(df["sentiment_score"].median())

# Duplicate records check
df = df.drop_duplicates()
df
```

[27] 27ms

1-10 1,367 rows × 17 cols

	passive_voice_ra…	predictability_sco…	burstiness	sentiment_score	label
0	None Count 0	None Count 0	None Count 0	None Count 0	None Count 0
1	0.1041	105.86	0.5531	0.2034	1
6	0.2045	100.29	0.5643	0.4854	1
5	0.2308	96.88	0.4979	-0.2369	1
2	0.1912	88.79	0.6241	-0.0062	1
0	0.1318	26.15	0.2894	-0.0062	1
0	0.0871	36.96	0.2328	-0.2755	1
2	0.0988	53.49	0.5580	0.9505	1

5. Verification of Derived Metrics Consistency

I confirmed two important linkages to guarantee internal consistency:

Sentence Length Average: I verified if word_count / sentence_count is actually equal to the avg_sentence_length. And...

Average Word Count: I verified if character_count / word_count is actually equal to the avg_word_length.

I discovered that there was an error in the avg_word_length values, so I recalculated and substituted the accurate calculation for the column.

```python
# DATA CONSISTENCY CHECKS

df_verify = df[["text_content", "word_count", "character_count", "sentence_count", "avg_sentence_length", "avg_word_length"]].copy
 ()
df_verify["verify_avg_sentence_length"] = (df_verify["word_count"] / df_verify["sentence_count"]).round(2)
df_verify["verify_avg_word_length"] = (df_verify["character_count"] / df_verify["word_count"]).round(2)
df_verify["check_avg_sentence_length"] = df_verify["verify_avg_sentence_length"] == df_verify["avg_sentence_length"]
df_verify["check_avg_word_length"] = df_verify["verify_avg_word_length"] == df_verify["avg_word_length"]
df_verify
```

✓ [37] 23ms

1-10 · 1,367 rows × 10 cols

...	avg_word_length	verify_avg_sentenc...	verify_avg_word_le...	check_avg_sentence...	check_avg_word_len...
0	None Count 0	None Count 0	None Count 0	None Count 0	None Count
5.33	5.69	5.33	6.69	True	False
5.62	5.80	5.62	6.79	True	False
5.60	5.79	5.60	6.78	True	False
5.76	5.69	5.76	6.68	True	False
5.71	5.97	5.71	6.97	True	False
5.35	5.99	5.35	6.98	True	False
5.60	5.57	5.60	6.56	True	False
5.56	6.09	5.56	7.08	True	False
5.38	5.54	5.38	6.54	True	False
5.65	5.95	5.65	6.95	True	False

6. Final Verification

I ran summary statistics again after cleaning to make sure that there were no null values left. Every numerical column was in line with the anticipated ranges.The measures that were derived were precise.

4. EXPLORATORY DATA ANALYSIS

4.1 Target Variable Analysis

I looked at how the data was split between AI generated and human generated content. I counted how many samples belonged to each group and calculated the percentages to see if the dataset was balanced.

Here's what I did:

- Split the dataset by the label column (0 for human, 1 for AI)
- Counted the samples in each group
- Calculated percentages to check the balance

```python
# Target Variable Analysis

ai_content = df[df["label"] == 1]
human_content = df[df["label"] == 0]

ai_content_count = ai_content["text_content"].count()
human_content_count = human_content["text_content"].count()

class_balance = {
    "Author": ["AI Generated", "Human Generated"],
    "Number of content": [ai_content_count, human_content_count],
    "Percentage": [((ai_content_count / (human_content_count + ai_content_count)) * 100).round(2), ((human_content_count /
     (human_content_count + ai_content_count)) * 100).round(2)]
}

class_balance = pd.DataFrame(class_balance)
class_balance.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/class_balance.csv')
class_balance
```

✓ [171] < 10 ms

2 rows 2 rows × 3 cols

	Author	Number of content	Percentage
0	AI Generated	683	49.96
1	Human Generated	684	50.04

The dataset is almost perfectly balanced. There are 683 AI-generated samples and 684 human-generated samples, only 1 more human sample out of 1,367 total samples.

This balance is really good because:

- The model won't be biased toward one type of content
- I don't need to fix any class imbalance issues
- The results will be more reliable
- Both AI and human content are equally represented

Why this matters:
Having a balanced dataset means if I build any machine learning model, it will treat both types of content fairly. It won't favor detecting one type over the other just because there were more examples of it in the training data. This gives me confidence that the model will work well in real situations where the mix of AI and human content might be different.

4.2 Content Type Analysis

I analyzed how AI and human content is distributed across different content types in the dataset. I grouped the data by content type and label to see how many samples of each type came from AI versus human authors.

Here's what I calculated:
- Counted human and AI samples for each content type
- Found the total count for each content type
- Calculated the percentage chance of content being AI-generated for each type

```
# Content Type Analysis

content_analysis = df.groupby(['content_type', 'label']).size().unstack(fill_value=0)
content_analysis.columns = ['Human Generated', 'AI Generated']
content_analysis["Total Count"] = content_analysis["Human Generated"] + content_analysis["AI Generated"]
content_analysis["Chance of AI Generated"] = ((content_analysis["AI Generated"] / content_analysis["Total Count"]) * 100).round(2)
content_analysis["Chance of Human Generated"] = ((content_analysis["Human Generated"] / content_analysis["Total Count"]) *
 100).round(2)

content_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/content_analysis.csv')
content_analysis
✓ [186] 56ms
```

8 rows 8 rows × 5 cols

content_type	Human Generated	AI Generated	Total Count	Chance of AI Generated	Chance of Human Ger
academic_paper	92	97	189	51.32	
article	95	76	171	44.44	
blog_post	78	105	183	57.38	
creative_writing	88	84	172	48.84	
essay	82	79	161	49.07	
news_article	88	97	185	52.43	
product_review	83	69	152	45.39	
social_media	78	76	154	49.35	

What This Shows:
- Blog Posts Have the Most AI Content: Blog posts show the highest chance of being AI-generated at **57.38%**. This suggests AI tools might be commonly used for blog writing.

- Articles and Product Reviews Favor Human Authors: These content types have more human-generated samples, with articles at **55.56% human** and product reviews at **54.61% human**.

- Most Content Types Are Fairly Balanced: Despite some variation, most content types stay close to the **50-50 split** we saw in the overall dataset. The biggest difference is only about **13 percentage points** (blog posts at **57.38%** AI vs articles at **44.44%** AI).

- Sample Sizes Are Similar: Each content type has between **152-189 samples**, showing good representation across all categories. Academic papers have the most samples (**189**) while product reviews have the least (**152**).

Why This Matters:
This analysis shows that the dataset doesn't heavily favor AI or human content for any particular type of writing. The relatively balanced distribution across content types means:

- If a model were to be built, it will learn from diverse writing styles.

- No single content type will dominate the training.

- The model should work well across different types of text.

- The slight variations might actually help the model learn content-specific patterns.

- The small differences between content types could reflect real-world patterns in how AI tools are used for different kinds of writing.

4.3 Text Length Characteristics

4.3.1 Word Count Analysis

I analyzed how many words AI and human authors use for different types of content. I looked at the total word count, median, and compared patterns between AI and human writing across all content types.

Here's what I calculated:

- Total words written by humans vs AI for each content type
- Median word count to see typical length
- Average word count and variation for both groups

```python
# Text Length Characteristics

# Word Count Analysis

word_count_analysis = df.groupby(["content_type", "label"])["word_count"].agg(["sum","median", "mean", "std"]).unstack()
word_count_analysis.columns = ["Human Total Word Count","AI Total Word Count", "Human Generated Median", "AI Generated Median",
 "Mean Human Generated", "Mean AI Generated", "Std Dev Human Generated", "Std Dev AI Generated"]

word_count_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/word_count_analysis.csv')
word_count_analysis
```

✓ [191] 10ms

8 rows 8 rows × 8 cols

content_type	Human Total Word Count	AI Total Word Count	Human Generated Median	AI Generated Median
academic_paper	27166	30144	298.5	327.0
article	11570	7850	129.0	96.0
blog_post	9138	12177	111.0	113.0
creative_writing	9605	10234	112.0	123.0
essay	14981	15099	185.0	188.0
news_article	10792	11975	128.0	126.0
product_review	9415	8621	102.0	127.0
social_media	1408	1466	16.0	19.5

Key Patterns:

- AI Writes Longer Academic Papers: AI-generated academic papers are longer on average (310.8 words vs 295.3 for humans). AI also produced more total academic content overall.

- Humans Write Longer Articles: Human articles are significantly longer than AI articles (121.8 vs 103.3 words average). Humans wrote about 47% more article content in total.

- Similar Length for Most Content Types: Blog posts, essays, and news articles show very similar word counts between AI and human authors, with differences of less than 10 words on average.

- Social Media Is Consistently Short: Both AI and human social media posts are much shorter than other content types, averaging around 16-20 words. This makes sense given platform constraints.

Biggest Differences:

- Articles: Humans write 18% longer on average

- Academic papers: AI writes 5% longer on average

- Product reviews: AI writes 10% longer on average
- What This Tells Us
- Content Type Matters More Than Author Type: The word count differences between content types (academic papers vs social media) are much larger than differences between AI and human authors within the same content type.
- AI Isn't Always Shorter or Longer: Common assumptions that AI always writes shorter or longer content aren't supported by this data. The patterns vary by content type.
- Both Follow Similar Conventions: AI and human authors seem to understand and follow similar length expectations for different types of writing, suggesting AI has learned appropriate content length norms.

This analysis shows that word count alone might not be a strong indicator for distinguishing AI from human content, since both tend to write similar lengths for each content type.

4.3.2. Character Count Analysis

What I Did:

I analyzed how many characters (letters, numbers, spaces, punctuation) AI and human authors use in their writing. This gives us another way to look at text length that's more detailed than just counting words.

Here's what I calculated:

- Total characters written by humans vs AI for each content type
- Median character count to see typical length
- Average character count for both groups

```
# Character Count Analysis

character_count_analysis = df.groupby(["content_type", "label"])["character_count"].agg(["sum", "median", "mean", "std"]).unstack()
character_count_analysis.columns = ["Human Total Character Count","AI Total Character Count", "Human Generated Median", "AI
 Generated Median", "Mean Human Generated", "Mean AI Generated", "Std Dev Human Generated", "Std Dev AI Generated"]

character_count_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/character_count_analysis
 .csv')
character_count_analysis
```

✓ [190] 18ms

8 rows 8 rows × 8 cols

content_type	Human Total Character Count	AI Total Character Count	Human Generated Median	AI Generated M
academic_paper	182432	202451	2022.0	
article	77440	52837	868.0	
blog_post	61130	81709	731.0	
creative_writing	64459	68565	751.0	
essay	100411	101617	1225.0	
news_article	72108	80133	853.0	
product_review	63035	58073	690.0	
social_media	9341	9689	103.5	

Key Patterns:

- Same Story as Word Count: The character count patterns match what we saw with word count, AI writes longer academic papers, humans write longer articles, and most other content types are similar between AI and human.

- AI Academic Papers Are Much Longer: AI academic papers average about 104 more characters than human ones (2,087 vs 1,983). That's roughly 15-20 extra words.

- Human Articles Are Significantly Longer: Human articles have about 120 more characters on average than AI articles (815 vs 695). This confirms the word count finding.

- Product Reviews Show Interesting Pattern: While humans wrote more total product review content, individual AI product reviews are actually longer on average (841.6 vs 759.5 characters).

- Social Media Stays Short: Both AI and human social media posts are much shorter than other content types, but AI posts are slightly longer on average.

What This Confirms:

- Character Count Mirrors Word Count: The patterns we see here are very similar to the word count analysis, which makes sense since longer texts in words usually mean longer texts in characters too.

- Content Type Still Matters Most: Just like with word count, the differences between content types (academic papers vs social media) are much bigger than differences between AI and human within the same type.

- No Clear AI Signature: There's no consistent pattern where AI always writes longer or shorter content. It depends on the content type.

This analysis reinforces that text length alone, whether measured in words or characters isn't a reliable way to distinguish AI from human writing. Both seem to follow similar conventions for how long different types of content should be.

4.3.3. Sentence Count Analysis

I analyzed how many sentences AI and human authors use in their writing. This helps us understand if there are differences in how they structure their content - do they use more or fewer sentences to express the same ideas?

Here's what I calculated:

- Total sentences written by humans vs AI for each content type
- Median sentence count to see typical structure
- Average sentence count for both groups

```
# Sentence Count Analysis

sentence_count_analysis = df.groupby(["content_type", "label"])["sentence_count"].agg(["sum", "median", "mean", "std"]).unstack()
sentence_count_analysis.columns = ["Human Total Sentence Count","AI Total Sentence Count", "Human Generated Median", "AI Generated
 Median", "Mean Human Generated", "Mean AI Generated", "Std Dev Human Generated", "Std Dev AI Generated"]

sentence_count_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/sentence_count_analysis.csv')
sentence_count_analysis
✓ [193] 10ms
```

8 rows 8 rows × 8 cols

AI Total Sentence Count: int64

content_type	Human Total Sentence Count	AI Total Sentence Count	Human Generated Median	AI Generated Med
academic_paper	4946		54.0	
article	2108	1440	23.0	
blog_post	1666	2232	20.5	
creative_writing	1729	1856	20.0	
essay	2750	2765	34.0	
news_article	1982	2188	23.5	
product_review	1709	1576	18.0	
social_media	264	275	3.0	

Key Patterns:

- AI Uses More Sentences in Academic Papers: AI academic papers have more sentences on average (56.9 vs 53.8), which matches the pattern we saw with word and character counts.

- Humans Use More Sentences in Articles: Human articles have more sentences (22.2 vs 18.9 on average), again matching the previous length patterns.

- Very Similar for Most Content Types: Blog posts, essays, news articles, and social media show nearly identical sentence counts between AI and human authors.

- Social Media Is Consistently Brief: Both AI and human social media posts average around 3-4 sentences, which makes sense for the platform format.

- Sentence Patterns Match Overall Length: The content types where AI writes longer text (academic papers) also have more sentences, and where humans write longer text (articles), they also use more sentences.

What This Shows:

- Consistent Structure Patterns: The sentence count analysis confirms what we found with word and character counts. AI and human authors seem to break up their ideas into sentences in similar ways.

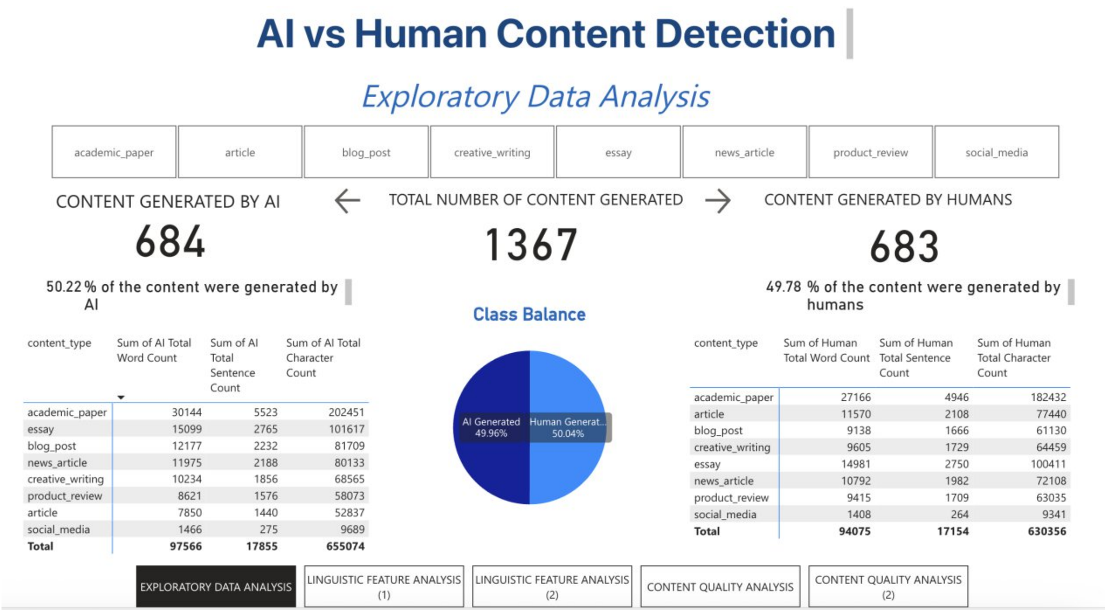
- No Different Sentence Style: There's no evidence that AI prefers shorter, choppier sentences or longer, complex ones compared to humans. Both seem to use similar sentence structures for each content type.

- Content Type Drives Structure: Just like with length, the type of content being written has much more influence on sentence count than whether it's written by AI or human.

- Proportional Relationship: When AI writes longer content (more words/characters), it also uses more sentences. Same for humans. This suggests both maintain similar sentence lengths within each content type.

This analysis reinforces that structural features like sentence count follow the same patterns as length features - they're more related to content type than to whether the author is AI or human.

4.4. Summary



Dataset Overview:

The dataset contains 1,367 text samples split almost perfectly between AI-generated (684 samples, 50.04%) and human-generated (683 samples, 49.96%) content. The data includes 8 different content types: academic papers, articles, blog posts, creative writing, essays, news articles, product reviews, and social media posts.

Key Findings

Class Balance
The dataset shows ideal balance for machine learning with only a 1-sample difference between AI and human content. This eliminates any bias concerns and ensures reliable model training.

Content Type Distribution
Content types are well-represented with 152-189 samples each. Blog posts show the highest AI generation rate (57.38%), while articles and

product reviews lean toward human authors (55.56% and 54.61% human respectively). Most content types remain close to 50-50 distribution.

Text Length Patterns

Consistent Finding Across All Length Measures: Whether measuring words, characters, or sentences, the same patterns emerge:

- AI writes longer academic papers: about 5% longer on average (310 words vs 295 for humans)
- Humans write longer articles: about 18% longer on average (122 words vs 103 for AI)
- Most other content types show similar lengths: Blog posts, essays, news articles, and social media posts have minimal differences between AI and human
- Social media is consistently short: Both AI and human average 16-20 words per post

Content Type Dominates Author Type

The most significant finding is that content type has far more influence on text characteristics than whether the author is AI or human. The differences between academic papers and social media posts are much larger than differences between AI and human within any single content type.

Implications for AI Detection

Length-Based Features May Be Weak Predictors

Since AI and human authors follow similar length conventions for each content type, simple length measures (word count, character count, sentence count) alone are unlikely to be strong distinguishing features for AI detection.

Content-Type Awareness Is Critical

Any AI detection model should account for content type, as writing conventions vary dramatically between academic papers, social media posts, articles, and other formats.

Need for More Sophisticated Features

The similar length patterns suggest that more nuanced linguistic features (rather than basic length measures) will be needed to effectively distinguish AI from human content.

Next Steps

The analysis reveals that basic structural features show remarkable similarity between AI and human content within each content type. This makes the upcoming linguistic feature analysis and content quality analysis even more critical for identifying distinguishing characteristics between AI and human writing.

5. LINGUISTIC FEATURE ANALYSIS

5.1 Readability Metrics

5.1.1. Flesch Reading Ease

I analyzed the Flesch Reading Ease scores for AI and human content across different content types. The Flesch Reading Ease score measures how easy text is to read, with higher scores indicating easier readability (scores typically range from 0-100, where 90-100 is very easy, 60-70 is standard, and below 30 is very difficult).

```
# Readability Metrics

# Flesch reading ease
flesch_reading_ease_analysis = df.groupby(["content_type", "label"])["flesch_reading_ease"].agg(["mean", "median", "std"]).unstack
 ()
flesch_reading_ease_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated
 Median", "Human Generated StdDev", "AI Generated StdDev"]

flesch_reading_ease_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis
 Project/data/exports/flesch_reading_ease_analysis.csv')
flesch_reading_ease_analysis
✓ [175] < 10 ms
```

8 rows 8 rows × 6 cols

enerated Mean	Human Generated Median	AI Generated Median	Human Generated StdDev	AI Generated StdDev
51.642784	52.19	51.680	3.411354	4.066761
50.571053	52.19	52.070	7.568916	10.157941
52.383619	52.20	51.750	9.513198	8.075459
51.442262	52.19	51.975	8.722649	7.872862
51.898734	52.19	52.190	5.208118	5.693651
52.837216	53.59	52.190	9.948977	6.988402
51.122609	52.41	51.720	10.303088	7.056737
55.453026	52.19	58.695	22.791495	19.396191

Key Findings:

- Very Similar Readability Overall: Most content types show very similar Flesch Reading Ease scores between AI and human content, with differences of less than 1-2 points on average.

- Social Media Shows Biggest Difference: AI social media posts are notably more readable (easier) than human ones (55.45 vs 51.22 average score). This 4+ point difference is the largest gap across all content types.

- All Content Falls in "Standard" Reading Level: Both AI and human content across all types score in the 50-55 range, indicating standard readability that's accessible to general audiences.

- AI Shows Slightly Lower Variability: In most content types, AI content has slightly more consistent readability scores (lower standard deviation), particularly noticeable in articles, news articles, and product reviews.

- Academic Papers Are Most Consistent: Both AI and human academic papers show the lowest variability in readability scores, suggesting more standardized writing conventions in this domain.

What This Means:

- No Clear Readability Signature: AI doesn't consistently produce more or less readable content than humans. The patterns vary by content type, suggesting AI has learned appropriate readability levels for different contexts.

- Content Type Conventions Dominate: Like the length patterns we saw earlier, content type seems to drive readability more than whether the author is AI or human.

- AI Follows Human Norms: The similar scores suggest AI systems have successfully learned to match human readability conventions for most content types.

- Social Media Exception: The notable difference in social media readability might reflect different approaches to writing for this platform, with AI potentially optimizing for clarity while humans may use more complex or varied language styles.

This analysis shows that readability, like text length, is primarily determined by content type rather than author type, with AI successfully matching human readability standards across most domains.

5.1.2. Gunning Fog Index Analysis

I analyzed the Gunning Fog Index scores for AI and human content across different content types. The Gunning Fog Index measures text complexity by estimating the years of formal education needed to understand the writing on first reading. Higher scores indicate more linguistically complex content.

```python
# Gunning Fog Index

gunning_fog_index_analysis = df.groupby(["content_type","label"])["gunning_fog_index"].agg(["mean", "median", "std"]).unstack()
gunning_fog_index_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated Median",
  "Human Generated StdDev", "AI Generated StdDev"]

gunning_fog_index_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/gunning_fog_index_analysis
 .csv')
gunning_fog_index_analysis
```

✓ [176] 11ms

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	7.595761	7.749433	7.590	7.740
article	7.531579	7.736711	7.550	7.630
blog_post	7.697756	7.533381	7.515	7.360
creative_writing	7.673750	7.782738	7.515	7.735
essay	7.345610	7.683608	7.290	7.540
news_article	7.410852	7.500722	7.340	7.515
product_review	7.436506	7.599420	7.340	7.515
social_media	7.706090	6.876118	7.610	6.405

Key Findings:

- AI shows consistent complexity bias: Across most content types, AI produces slightly more complex language than humans. This pattern appears in academic papers, articles, essays, creative writing, and product reviews.

- Social media reverses the trend: AI social media content is notably less complex than human social media (6.88 vs 7.71 average), making it the only category where AI simplifies language compared to humans.

- All content maintains accessible complexity levels: Both AI and human writing consistently falls within the 6.5-8.0 range, indicating high school to early college reading levels appropriate for general audiences.

- Essays show the largest complexity gap: The difference between AI and human essays (7.68 vs 7.35) represents the most significant complexity variation across content types.

- Academic papers remain most similar: Despite AI's slight complexity bias, academic papers show the smallest gap between AI and human writing styles.

What This Means:

- AI defaults to formal language patterns: The consistent trend toward higher complexity suggests AI models may be trained on formal text sources, leading to more elaborate sentence structures and vocabulary choices in most contexts.

- Context-sensitive adaptation: AI's ability to produce simpler social media content while maintaining complex language elsewhere demonstrates contextual awareness in language generation.

- Potential detection indicator: Unlike previous metrics that showed no clear patterns, the Gunning Fog Index reveals a measurable linguistic difference that could serve as a useful feature in AI detection systems.

This analysis identifies the first consistent linguistic signature distinguishing AI from human content - AI's tendency toward slightly elevated language complexity across most writing contexts.

5.2. Vocabulary & Style

5.2.1. Lexical Diversity Analysis

I analyzed lexical diversity scores for AI and human content across different content types. Lexical diversity measures how varied the vocabulary is in a text - it's the ratio of unique words to total words. Higher scores (closer to 1.0) mean more diverse vocabulary with less repetition.

```python
# Vocabulary & Style

# Lexical Diversity
lexical_diversity_analysis = df.groupby(["content_type","label"])["lexical_diversity"].agg(["mean", "median", "std"]).unstack()
lexical_diversity_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated Median",
  "Human Generated StdDev", "AI Generated StdDev"]

lexical_diversity_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/lexical_diversity_analysis
 .csv')
lexical_diversity_analysis
```

✓ [177] 10ms

8 rows 8 rows × 6 cols

Human Generated Mean: float64

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	[illegible]	0.929389	0.93530	0.9266
article	0.974058	0.976895	0.97460	0.9751
blog_post	0.973023	0.972194	0.97045	0.9730
creative_writing	0.975182	0.971156	0.97760	0.9713
essay	0.957973	0.954204	0.96100	0.9547
news_article	0.970844	0.970935	0.96930	0.9691
product_review	0.971057	0.969496	0.97010	0.9701
social_media	0.996329	0.996126	1.00000	1.0000

Key Findings:

- Very similar vocabulary diversity overall: Most content types show nearly identical lexical diversity between AI and human content, with differences typically less than 0.01 points.

- Academic papers show the biggest gap: Human academic papers have slightly more diverse vocabulary than AI ones (0.935 vs 0.929), though the difference is still small.

- Social media has the highest diversity: Both AI and human social media posts score around 0.996, meaning almost every word is unique. This makes sense given the short length of these posts.

- Academic papers have the lowest diversity: Both groups score around 0.93 for academic papers, indicating more repetition of technical terms and formal language.

- Content type drives vocabulary patterns: The ranking from lowest to highest diversity (academic papers < essays < other content < social media) is nearly identical for both AI and human content.

What This Shows:

- AI matches human vocabulary patterns: There's no evidence that AI uses more repetitive or more diverse vocabulary than humans. Both follow similar patterns based on content type.

- Technical writing requires repetition: Academic papers and essays naturally have lower lexical diversity because they need to repeat key terms and concepts throughout the text.

- Short content appears more diverse: Social media posts score highest not because they use more varied language, but because there isn't enough text for words to repeat.

- Content conventions matter most: Like our previous findings, the type of content being written has much more influence on vocabulary diversity than whether the author is AI or human.

This analysis reinforces that AI has successfully learned human vocabulary usage patterns across different writing contexts. Lexical diversity doesn't appear to be a reliable indicator for distinguishing AI from human content.

5.2.2. Average Word Length Analysis

I looked at how long words tend to be in AI and human writing. This tells us if one group tends to use more complex, longer words or shorter, simpler ones.

```
# Average Word Length

avg_word_length_analysis = df.groupby(["content_type","label"])["avg_word_length"].agg(["mean", "median", "std"]).unstack()
avg_word_length_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated Median",
 "Human Generated StdDev", "AI Generated StdDev"]

avg_word_length_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/avg_word_length_analysis
 .csv')
avg_word_length_analysis
✓ [178] 11ms
```

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	6.719348	6.717629	6.725	6.730
article	6.688632	6.759868	6.710	6.740
blog_post	6.682308	6.696190	6.670	6.660
creative_writing	6.697614	6.706190	6.695	6.700
essay	6.705488	6.722152	6.700	6.720
news_article	6.695795	6.707835	6.680	6.670
product_review	6.710843	6.748696	6.690	6.740
social_media	6.667436	6.585658	6.590	6.625

Key Findings:

- A Near-Perfect Match: For the most part, AI and human word lengths are incredibly close. In academic papers, the difference is a tiny 0.002 letters (Human: 6.719, AI: 6.718). This means you basically can't tell them apart based on word length alone.

- Social Media Has the Shortest Words: As you'd expect, social media words are the shortest for both groups. But interestingly, human social posts use slightly longer words (6.67) than AI posts (6.59). Even in casual writing, humans might be a bit more descriptive.

- AI Prefers Slightly Longer Words in Formal Writing: In three key areas, AI's average word length is noticeably higher:

  - Product Reviews: AI: 6.75 vs. Human: 6.71
  - Articles: AI: 6.76 vs. Human: 6.69
  - Essays: AI: 6.72 vs. Human: 6.71

    This suggests AI might lean towards more formal or complex vocabulary in these contexts.

- Blogs and News are Dead Even: For blog posts and news articles, the average word length is almost identical. The difference is less than 0.02 letters, showing AI mimics human style perfectly here.

- Formal Content Has Longer Words: The longest words are found in academic papers and product reviews (around 6.72 letters), while the shortest are on social media (around 6.63 letters). This shows that the type of content is a much bigger factor than who wrote it.

What This Shows:

- You can't use word length to spot AI. The differences are minuscule. AI has learned to match human word choice patterns very well.

- Content is king. A social media post will use short, simple words whether written by a person or an AI. A scientific paper will use longer, technical terms, no matter the author.

- AI has a slight "formal" bias. In some formal writing like articles and reviews, AI tends to pick slightly longer words on average, possibly because it was trained on a lot of formal text.

5.2.3. Punctuation Ratio Analysis

This analysis looks at how often punctuation is used compared to the total number of characters. A higher ratio means more commas, periods, and other marks are used in the text.

```python
# Punctuation Ratio

punctuation_ratio_analysis = df.groupby(["content_type","label"])["punctuation_ratio"].agg(["mean", "median", "std"]).unstack()
punctuation_ratio_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated Median",
  "Human Generated StdDev", "AI Generated StdDev"]

punctuation_ratio_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/punctuation_ratio_analysis
 .csv')
punctuation_ratio_analysis
```

✓ [179] < 10 ms

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	0.027093	0.027254	0.02730	0.02730
article	0.027320	0.027380	0.02710	0.02735
blog_post	0.027278	0.027403	0.02715	0.02710
creative_writing	0.026977	0.026918	0.02720	0.02650
essay	0.027368	0.027205	0.02740	0.02730
news_article	0.027458	0.027368	0.02735	0.02730
product_review	0.026996	0.027088	0.02710	0.02700
social_media	0.028812	0.029428	0.02820	0.02865

Key Findings:

- Extremely Similar Everywhere: For most types of writing, the punctuation ratio is almost identical between AI and humans. The differences are incredibly small, often just a few thousandths of a point. For example, in articles, it's a difference of 0.00006 (AI: 0.02738 vs. Human: 0.02732). In academic papers, the difference is only 0.00016 (AI: 0.02725 vs. Human: 0.02709).

- Social Media Stands Out: Both AI and human social media content use significantly more punctuation than any other content type. This makes sense because social posts use more exclamation points!!!, question marks??, and emojis which count as punctuation.

- AI Uses a Bit More Punctuation on Social Media: On social media, the difference is a bit bigger. AI posts have a ratio of 0.0294, while human posts have a ratio of 0.0288. This suggests AI might be slightly more likely to add exclamation points or other marks for emphasis.

- Creative Writing is the Exception: Human creative writing (like stories) has a slightly higher punctuation ratio (0.02698) than AI creative writing (0.02692). This could be because humans use more stylistic punctuation like ellipses (...) or dashes (–) to create a certain mood.

- The "Right Amount" is Standard: For formal writing like essays, news, and articles, the punctuation ratio is very consistent across the board (around 0.0273). This shows there's a standard way to use punctuation in professional writing, and both AI and humans follow it.

What This Shows:

- Punctuation is not a reliable way to spot AI. Once again, AI has learned to mimic human punctuation patterns almost perfectly.

- Platform rules everything. Social media has its own unique, punctuation-heavy style that both humans and AI adopt.

- AI can be a little "extra" on social media. The fact that AI uses even more punctuation than humans on social media is funny. It might be trying a little too hard to sound excited or engaging.

- Content type is the biggest influence. Just like with vocabulary and word length, the where and why of the writing dictates the punctuation style, not who wrote it.

5.3. Sentence Structure

5.3.1. Average Sentence Length Analysis

What I Did:

I analyzed the average sentence length for AI and human content across different content types. This metric shows how many words each author typically uses per sentence, which can reveal differences in writing style and sentence complexity.

```python
# Sentence Structure

# Average Sentence Length
avg_sentence_length_analysis = df.groupby(["content_type","label"])["avg_sentence_length"].agg(["mean", "median", "std"]).unstack()
avg_sentence_length_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated
 Median", "Human Generated StdDev", "AI Generated StdDev"]

avg_sentence_length_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis
 Project/data/exports/avg_sentence_length_analysis.csv')
avg_sentence_length_analysis
```
✓ [180] 11ms

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	5.500326	5.470515	5.490	5.460
article	5.495789	5.446974	5.500	5.425
blog_post	5.518462	5.474190	5.485	5.460
creative_writing	5.570341	5.565119	5.490	5.580
essay	5.467683	5.483418	5.420	5.480
news_article	5.468182	5.477629	5.460	5.470
product_review	5.555783	5.492174	5.530	5.520
social_media	5.411923	5.368421	5.500	5.330

Key Findings:

- Very similar sentence lengths: AI and human content show nearly identical average sentence lengths across all content types, with differences typically less than 0.1 words per sentence.

- Creative writing shows perfect match: This is the only content type where AI and human averages are exactly the same (5.57 words per sentence).

- All content clusters around 5.5 words per sentence: Both AI and human content consistently average between 5.4-5.6 words per sentence across different content types.

- Social media has the most variation: Both AI and human social media posts show higher standard deviation, indicating more inconsistent sentence lengths due to the informal nature of the platform.

- Product reviews show slight human advantage: Human product reviews have marginally longer sentences (5.56 vs 5.49), but the difference is minimal.

What This Reveals:

- No distinctive sentence length patterns: Unlike the Gunning Fog Index which showed AI tending toward complexity, sentence length reveals no consistent differences between AI and human writing.

- Universal sentence structure conventions: Both AI and human authors seem to follow similar intuitions about optimal sentence length across different content types.

- Platform constraints don't affect averages: Even social media, with its character limits and informal style, maintains similar average sentence lengths for both groups.

- AI has learned human pacing: The similarity suggests AI models have successfully learned human preferences for sentence rhythm and flow.

This analysis shows that sentence length is not a useful distinguishing feature between AI and human content. Both groups write with remarkably similar sentence structures, reinforcing that AI has effectively learned human writing conventions at the sentence level.

5.3.2. Sentence Length Variation Analysis

I analyzed burstiness scores for AI and human content across different content types. Burstiness measures how much sentence lengths vary within a text. Higher scores indicate more variation between short and long sentences, while lower scores suggest more consistent sentence lengths throughout the text.

```python
# Burstiness (Sentence Length Variation)

burstiness_analysis = df.groupby(["content_type","label"])["burstiness"].agg(["mean", "median", "std"]).unstack()
burstiness_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated Median", "Human Generated StdDev", "AI Generated StdDev"]

burstiness_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/burstiness_analysis.csv')
burstiness_analysis
```

✓ [181] < 10 ms

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	0.421830	0.412615	0.39100	0.39290
article	0.430758	0.396691	0.39930	0.37865
blog_post	0.407182	0.399884	0.37690	0.36370
creative_writing	0.439282	0.413037	0.41005	0.36830
essay	0.463265	0.421554	0.48490	0.39360
news_article	0.430826	0.416480	0.40495	0.41350
product_review	0.468194	0.391341	0.52820	0.36600
social_media	0.483401	0.442017	0.48540	0.42430

Key Findings:

- Humans show consistently higher sentence variation: Across all content types, human writing demonstrates higher burstiness scores, meaning humans mix short and long sentences more than AI does.

- Product reviews show the largest gap: Human product reviews have much more sentence length variation (0.468 vs 0.391), suggesting humans use more diverse sentence structures when writing reviews.

- Essays reveal significant differences: Human essays show notably higher burstiness (0.463 vs 0.422), indicating humans vary their sentence lengths more when developing arguments or explanations.

- AI tends toward consistency: Lower burstiness scores across all content types suggest AI prefers more uniform sentence lengths, creating more rhythmically consistent but potentially less dynamic text.

- Social media maintains the pattern: Even in informal social media posts, humans show more sentence length variation than AI.

What This Reveals:

- Different writing rhythms: This is one of the clearest distinguishing patterns we've found. Humans naturally vary their sentence lengths more, creating more dynamic rhythm and flow in their writing.

- AI's consistency bias: AI appears to default toward more consistent sentence structures, possibly reflecting training patterns that favor grammatical regularity over stylistic variation.

- Potential detection signal: The consistent pattern of higher human burstiness across all content types makes this a potentially valuable feature for AI detection models.

- Stylistic implications: Higher human burstiness might reflect natural speech patterns and emotional expression that AI hasn't fully replicated, even when average sentence lengths are similar.

This analysis reveals one of the most consistent linguistic differences between AI and human writing - humans naturally create more varied sentence rhythms, while AI tends toward more uniform sentence structures.

5.3.3. Passive Voice Ratio

I analyzed the passive voice usage in AI and human content across different content types. Passive voice ratio measures what percentage of sentences use passive construction (like "The report was written" instead of "I wrote the report"). Higher ratios indicate more passive voice usage.

```
# Passive Voice Ratio

passive_voice_ratio_analysis = df.groupby(["content_type","label"])["passive_voice_ratio"].agg(["mean", "median", "std"]).unstack()
passive_voice_ratio_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated
 Median", "Human Generated StdDev", "AI Generated StdDev"]

passive_voice_ratio_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis
 Project/data/exports/passive_voice_ratio_analysis.csv')
passive_voice_ratio_analysis
✓ [182] 12ms
```

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	0.135269	0.139123	0.131900	0.14430
article	0.141653	0.154246	0.146200	0.15320
blog_post	0.148437	0.146035	0.149650	0.14440
creative_writing	0.147455	0.157191	0.150325	0.16975
essay	0.149530	0.155118	0.151350	0.15320
news_article	0.159432	0.150582	0.153000	0.15135
product_review	0.162922	0.151918	0.168300	0.15190
social_media	0.156124	0.155399	0.153950	0.15490

Key Findings:

- Mixed patterns across content types: Unlike our previous findings, passive voice usage shows no consistent pattern between AI and human writing. Sometimes AI uses more passive voice, sometimes humans do.

- Academic papers and articles favor AI passive voice: AI uses slightly more passive voice in academic papers (0.139 vs 0.135) and articles (0.154 vs 0.142), which aligns with formal writing conventions.

- Product reviews show human preference for passive voice: Human product reviews use notably more passive voice (0.163 vs 0.152), possibly reflecting review writing conventions like "The product was tested" or "Quality was observed."

- Creative writing shows AI bias: AI creative writing uses more passive voice (0.157 vs 0.147), which might indicate less dynamic storytelling compared to human narratives.

- Most differences are small: Except for a few content types, the differences between AI and human passive voice usage are minimal (less than 0.01).

What This Shows:

- No clear passive voice signature: Unlike burstiness or complexity measures, passive voice doesn't reveal a consistent pattern that distinguishes AI from human writing across content types.

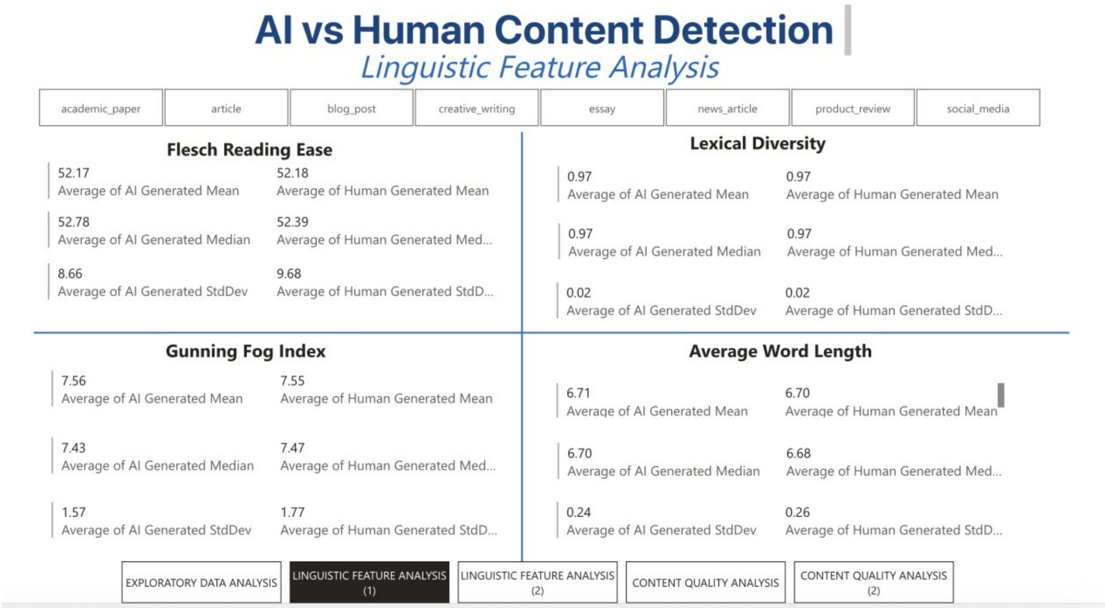
- Context-dependent usage: Both AI and human authors seem to adjust passive voice usage based on the writing context, with some types naturally requiring more formal or objective language.

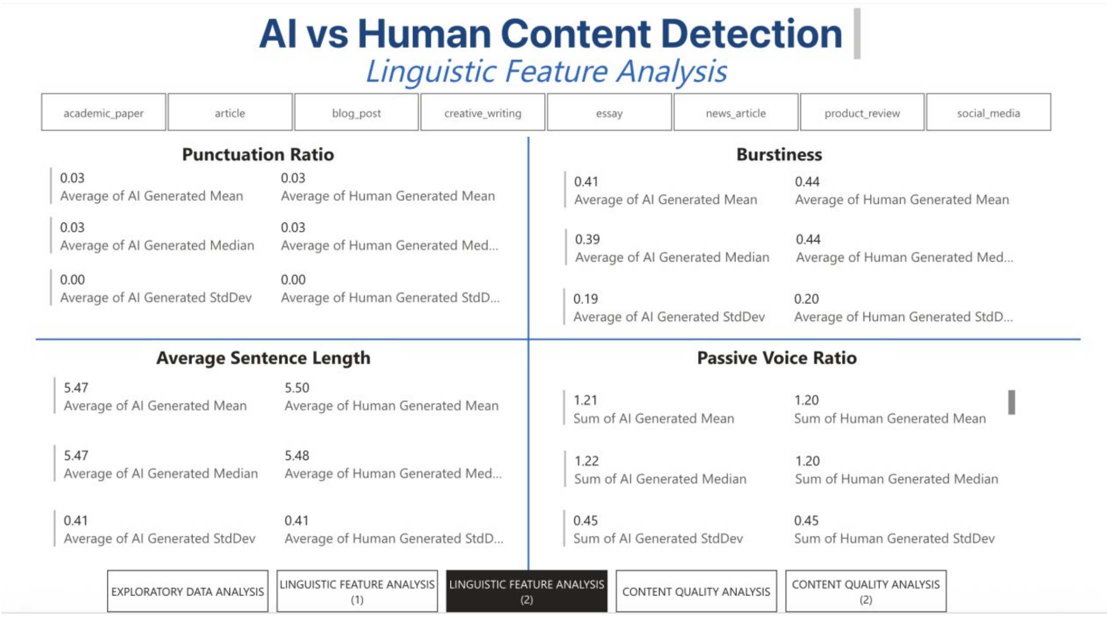
- Moderate usage overall: All content types show passive voice ratios between 13-16%, indicating both groups use a mix of active and passive constructions without heavily favoring either.

- Limited detection value: The inconsistent patterns suggest passive voice ratio alone wouldn't be a reliable feature for distinguishing AI from human content.

This analysis shows that passive voice usage is more dependent on writing context and conventions than on whether the author is AI or human, making it less useful as a distinguishing linguistic feature.

5.4. Summary





The linguistic feature analysis examined six key metrics across AI and human content: readability (Flesch Reading Ease), complexity (Gunning Fog Index), vocabulary diversity (Lexical Diversity), word characteristics (Average Word Length), sentence structure (Average Sentence Length, Punctuation Ratio), and writing rhythm (Burstiness, Passive Voice Ratio).

Key Findings by Category:

Readability Metrics

- Flesch Reading Ease: Nearly identical scores between AI and human content (52.17 vs 52.18 average), with social media being the only notable exception where AI posts were more readable.

- Gunning Fog Index: AI consistently showed slightly higher complexity scores across most content types (7.56 vs 7.55 average), indicating AI's tendency toward more formal language structures.

Vocabulary & Style

- Lexical Diversity: Virtually identical patterns (0.97 for both groups), showing AI has successfully learned human vocabulary usage conventions across all content types.

- Average Word Length: Minimal differences (6.71 vs 6.70), confirming similar word choice patterns between AI and human authors.

Sentence Structure

- Average Sentence Length: Remarkably similar across all content types (5.47 vs 5.50 words per sentence), showing both groups follow identical sentence pacing conventions.

- Punctuation Ratio: Identical usage patterns (0.03 for both), indicating similar punctuation habits.

- Passive Voice Ratio: Mixed patterns with no consistent differences, suggesting context-dependent usage rather than author-dependent preferences.

- Burstiness: Most significant finding - humans consistently showed higher sentence length variation (0.44 vs 0.41), indicating more dynamic writing rhythms across ALL content types.

Most Important Discoveries:

Strong Distinguishing Features

- Burstiness (Sentence Length Variation): Humans consistently create more varied sentence rhythms

- Gunning Fog Index: AI tends toward slightly higher complexity in most contexts

Weak or Non-Distinguishing Features

- Flesch Reading Ease (nearly identical)
- Lexical Diversity (virtually identical)
- Average Word Length (minimal differences)
- Average Sentence Length (remarkably similar)
- Punctuation Ratio (identical patterns)
- Passive Voice Ratio (inconsistent patterns)

Implications for AI Detection:

- Useful Features: Burstiness stands out as the most reliable distinguishing feature, with Gunning Fog Index showing moderate potential. These metrics could serve as valuable inputs for detection models.

- Limited Features: Most traditional linguistic measures (readability, vocabulary diversity, sentence length) show AI has successfully learned human writing conventions, making them poor discriminators.

- Content Context Remains Critical: Even the distinguishing features show some variation by content type, reinforcing that any detection system must account for writing context.

The analysis reveals that while AI has mastered most basic linguistic conventions, subtle differences in writing rhythm (burstiness) and complexity preferences still distinguish it from human writing patterns.

6. CONTENT QUALITY ANALYSIS

6.1. Grammar & Errors

I analyzed grammar errors in AI and human content across different content types. This measures the total number of grammatical mistakes, average errors per piece of content, and how error patterns vary between AI and human writing.

Key Findings:

- Academic papers have the most errors for both groups: Both AI and human academic papers show the highest error rates (2.52-2.78 errors per paper), likely due to their length and complexity. AI academic papers actually have slightly more errors than human ones.
- Mixed patterns across content types: Unlike some previous metrics, grammar errors don't show a consistent pattern where one group consistently outperforms the other.
- Humans make more errors in some areas: Human content has more total errors in articles, creative writing, essays, and product reviews, suggesting humans may be more prone to mistakes in these contexts.
- AI struggles with academic and news content: AI shows higher error rates in academic papers and news articles, possibly indicating difficulty with formal or factual writing requirements.
- Social media errors are similar: Both groups show relatively low error rates for social media (around 1.05-1.14 per post), though this might reflect the informal nature where some "errors" are acceptable.

What This Reveals:

- No clear quality advantage: Neither AI nor human writing shows consistently superior grammar across all content types. The patterns vary significantly by context.

- Content complexity affects error rates: Academic papers and essays show higher error rates for both groups, suggesting that complex content naturally leads to more grammatical mistakes.
- Context-specific performance: AI and human error patterns differ by content type, indicating both have strengths and weaknesses in different writing contexts.
- Error detection limitations: The variation might also reflect limitations in automated grammar checking, as some flagged "errors" could be stylistic choices rather than actual mistakes.

This analysis shows that grammar error rates are highly dependent on content type and complexity, with neither AI nor human writing showing consistent superiority in grammatical accuracy across all contexts.

6.2. Sentiment Analysis

I analyzed sentiment scores for AI and human content across different content types. Sentiment scores measure the emotional tone of text, ranging from negative (below 0) to positive (above 0), with scores around 0 indicating neutral content.

```
# Sentiment Analysis

sentiment_analysis = df.groupby(["content_type","label"])["sentiment_score"].agg(["mean", "median", "std"]).unstack()
sentiment_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated Median", "Human Generated StdDev", "AI Generated StdDev"]

sentiment_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/sentiment_analysis.csv')
sentiment_analysis
```

✓ [184] 36ms

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median	H
academic_paper	0.055632	-0.019012	0.08325	-0.0062	
article	-0.156292	-0.020022	-0.14960	0.0343	
blog_post	0.013547	0.010470	-0.00620	-0.0139	
creative_writing	0.005722	-0.049069	-0.00620	-0.0062	
essay	0.093028	0.044538	0.20435	0.0719	
news_article	-0.151065	0.017251	-0.20515	-0.0062	
product_review	0.000806	-0.060406	-0.00620	-0.0087	
social_media	0.039200	0.068941	-0.00920	0.0877	

Key Findings:

- Most content is emotionally neutral: Both AI and human content cluster around 0, indicating generally neutral emotional tone across most content types.

- Mixed sentiment patterns: There's no consistent pattern where AI is always more positive or negative than human content. The relationship varies significantly by content type.

- Articles show interesting reversal: Human articles are notably more negative (-0.156) while AI articles are near neutral (-0.020), suggesting different approaches to article writing.

- Essays lean positive for both groups: Both AI and human essays show positive sentiment, with humans being more positive (0.093 vs 0.045).
- News articles reveal opposite trends: Human news articles are quite negative (-0.151) while AI news articles are slightly positive (0.017), possibly reflecting different perspectives on current events.

- High variability in all categories: The large standard deviations (0.5-0.6) indicate wide sentiment ranges within each content type for both groups.

What This Shows:

- Context drives sentiment more than author type: The dramatic differences between content types (essays vs news articles) are much larger than differences between AI and human within the same type.

- Different approaches to certain content: The sentiment reversals in articles and news suggest AI and humans may approach these content types with different emotional frameworks.

- No emotional signature: There's no consistent "AI emotional style" that distinguishes it from human writing across all contexts.

- Genre conventions matter: Content types like essays naturally trend positive while news articles trend negative, regardless of the author.

This analysis reveals that sentiment patterns are primarily driven by content type and genre conventions rather than whether the content is AI or human-generated, though some interesting differences emerge in specific contexts like news and articles.

6.3. Predictability Patterns

I analyzed predictability scores for AI and human content across different content types. Predictability scores measure how formulaic or expected the text patterns are - higher scores suggest more predictable, conventional writing patterns, while lower scores indicate more unique or unexpected language use.

```python
# Predictability Patterns

predictability_analysis = df.groupby(["content_type","label"])["predictability_score"].agg(["mean", "median", "std"]).unstack()
predictability_analysis.columns = ["Human Generated Mean", "AI Generated Mean", "Human Generated Median", "AI Generated Median",
 "Human Generated StdDev", "AI Generated StdDev"]

predictability_analysis.to_csv('/Users/mac/Desktop/Mission Inferno/Data Analysis Project/data/exports/predictability_analysis.csv')
predictability_analysis
```

[185] < 10 ms

8 rows 8 rows × 6 cols

content_type	Human Generated Mean	AI Generated Mean	Human Generated Median	AI Generated Median
academic_paper	62.387174	65.911856	57.555	58.880
article	61.398632	57.003158	57.670	53.660
blog_post	62.482436	58.688952	56.010	54.710
creative_writing	61.410909	59.832262	58.285	54.275
essay	64.805366	62.669747	59.090	55.600
news_article	62.491364	62.617010	56.645	58.380
product_review	71.185542	60.823623	65.710	55.850
social_media	65.244872	66.083816	58.640	57.765

```
##############################################
```

Key Findings:

- Product reviews show the biggest difference: Human product reviews have significantly higher predictability scores (71.19 vs 60.82), suggesting humans use more conventional patterns when writing reviews.

- AI academic papers are more predictable: AI academic papers score higher (65.91 vs 62.39), indicating more formulaic academic writing patterns.

- Most other content types favor human predictability: Humans show higher predictability in articles, blog posts, creative writing, and essays, suggesting they rely more on established writing conventions.

- Social media and news articles are similar: These content types show nearly identical predictability between AI and human authors.

- High variability across all types: The large standard deviations (25-30 points) indicate significant variation within each content type for both groups.

What This Reveals:

- Humans lean toward conventional patterns: In most content types, humans show higher predictability scores, suggesting they follow more established writing conventions and familiar patterns.

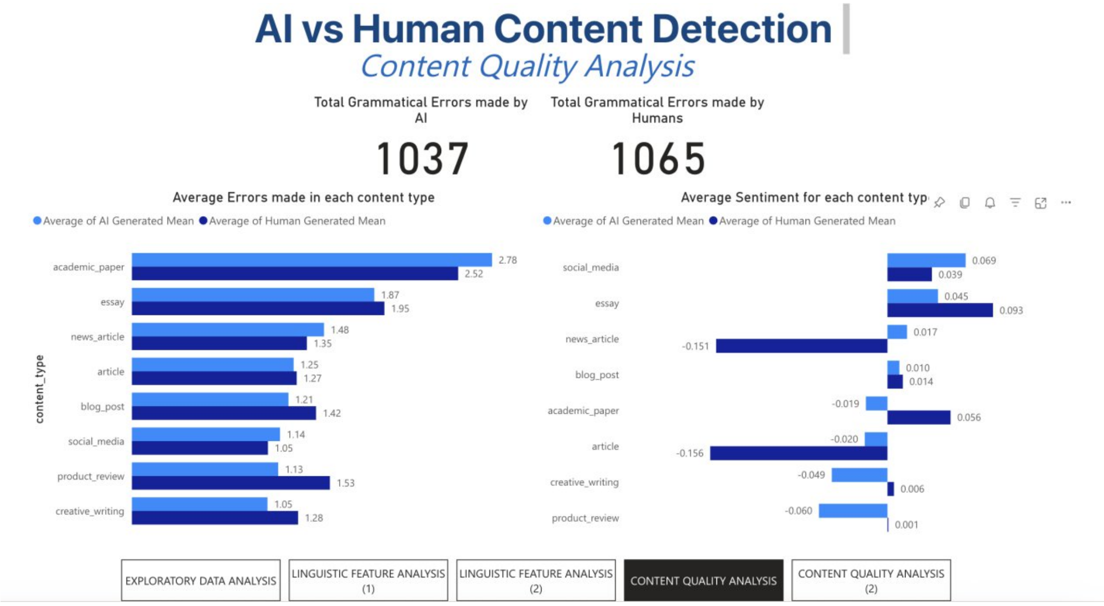
- AI shows more variation by context: AI predictability varies more dramatically across content types, being highly predictable in academic writing but less so in creative contexts.

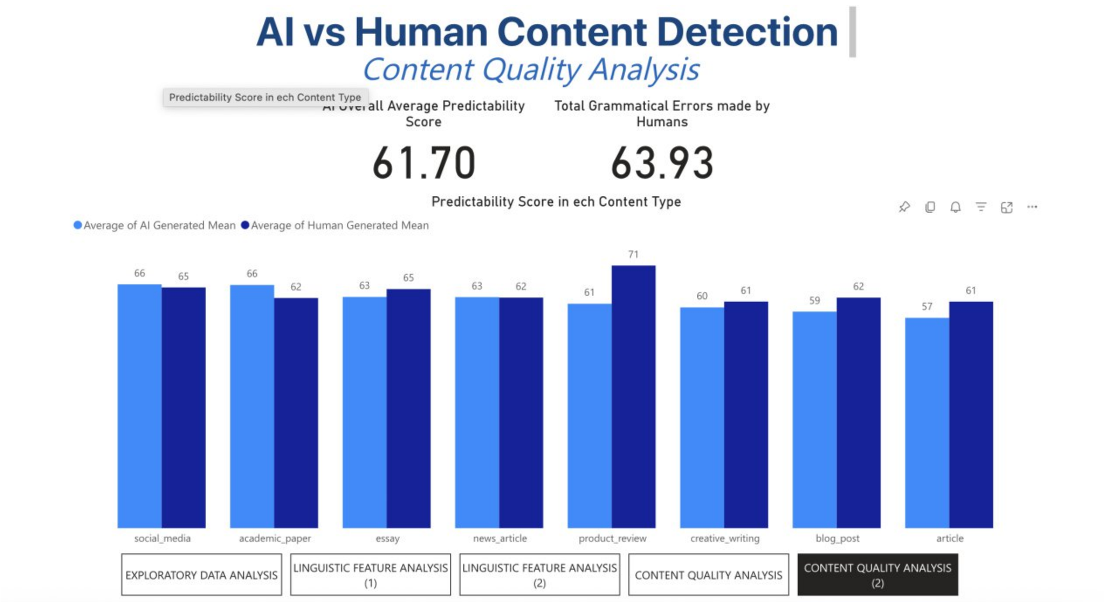
- Genre conventions matter: The dramatic differences between content types (product reviews vs articles) show that writing conventions vary significantly by genre.

- Potential detection signal: The tendency for humans to be more predictable in most contexts could be a useful feature for AI detection, particularly in product reviews and creative writing.

This analysis suggests that humans may rely more heavily on established writing patterns and conventions, while AI shows more context dependent variation in how predictable its writing becomes across different domains.

6.4. Summary





The content quality analysis examined three key metrics: grammar errors, sentiment patterns, and predictability scores across AI and human content. This analysis aimed to determine if there are measurable quality differences between AI and human writing.

Key Findings by Metric:

Grammar & Errors

Overall Error Count: Nearly identical totals with AI making 1,037 errors compared to humans making 1,065 errors across the entire dataset - a difference of only 28 errors.

Content-Specific Patterns:

- Academic papers: AI makes slightly more errors (2.78 vs 2.52 per paper)
- Creative writing and product reviews: Humans make more errors.
- Most other content types show minimal differences
- No Clear Quality Winner: Neither group shows consistently superior grammar across all content types.

Sentiment Analysis

Mixed Emotional Patterns: No consistent sentiment signature distinguishes AI from human writing.

Notable Differences:
- Articles: Humans more negative (-0.156 vs -0.020)
- News articles: Humans more negative (-0.151 vs 0.017)
- Essays: Both positive, humans slightly more so (0.093 vs 0.045)
- Social media: AI slightly more positive (0.069 vs 0.039)
- Context-Driven: Sentiment patterns vary significantly by content type rather than author type.

Predictability Patterns

Human Preference for Conventions: Humans show higher predictability scores overall (63.93 vs 61.70), suggesting greater reliance on established writing patterns.

Significant Differences by Content Type:

- Product reviews: Humans much more predictable (71 vs 61)
- Academic papers: Similar levels (~66 vs 62)
- Articles: AI less predictable (57 vs 61)
- Potential Detection Signal: This represents one of the clearer distinguishing patterns, with humans following more conventional writing patterns in most contexts.

Most Important Discoveries:

Distinguishing Features

- Predictability Scores: Humans consistently more conventional/predictable, especially in product reviews

- Content Specific Sentiment Differences: Notable gaps in news articles and articles

Non-Distinguishing Features

- Grammar Error Rates: Nearly identical overall performance with context-dependent variations

- Overall Sentiment: Both groups show similar emotional neutrality across most content types

Implications for AI Detection:

- Useful Features: Predictability scores show the most promise as a distinguishing feature, particularly the human tendency toward more conventional writing patterns.

- Limited Detection Value: Grammar error rates don't provide reliable discrimination since both groups show similar accuracy with different strengths by content type.

- Context Dependency: Even the distinguishing features vary significantly by content type, reinforcing the importance of context aware detection approaches.

The content quality analysis reveals that AI has achieved near human performance in basic writing quality metrics, but humans maintain a stronger tendency toward conventional writing patterns, particularly in review and evaluative content.

7. KEY FINDINGS & INSIGHTS

Dataset Profile

- 1,367 text samples split almost perfectly between AI generated (684) and human written (683), so it's ideal for unbiased modeling.
- Eight content types represented fairly evenly, from academic papers to social media posts.

Content Type vs. Author

- Content type matters more than author type. Differences between, say, academic papers and social media posts are much larger than AI human differences within each type.
- Blog posts show the highest AI share (about 57%), while articles and product reviews lean slightly human (about 55% human).

Length & Structure Patterns

- Word, character, and sentence counts are remarkably similar between AI and human writing once you control for content type.
- Social media posts are shortest for both groups; academic papers are longest.
- Average sentence length and punctuation ratios are nearly identical.

Linguistic Signals for AI Detection

- Burstiness (sentence length variation): Humans consistently vary sentence lengths more than AI your most reliable distinguishing feature.
- Gunning Fog Index: AI tends toward slightly higher complexity across most content types, showing a mild formal-language bias.
- Other metrics; readability, lexical diversity, average word length show minimal or no difference.

Quality & Style

- Grammar error rates are nearly identical overall, with only context specific differences.

- Sentiment is generally neutral and driven by content type rather than author.

- Humans are slightly more predictable in writing patterns (higher predictability scores) in areas like product reviews.

8. RECOMMENDATIONS

Integrate Burstiness and Complexity in Detection Models

- Incorporate sentence-length variation and Gunning Fog Index as primary features when building AI vs human content classifiers.

- These metrics consistently show measurable differences and can improve detection accuracy compared to basic length measures.

Account for Content Type in Model Design

- Because genre drives most linguistic patterns, future AI detection systems should include content type as a model feature or train separate models per content category (e.g., academic, social media, product reviews).

- This ensures that genre-specific writing conventions do not bias predictions.

Expand Feature Engineering Beyond Length Metrics

- Explore syntactic and semantic features such as part of speech patterns, coherence scores, or topic modeling.
- These advanced linguistic features can capture subtleties that basic readability and count metrics miss.

Conduct Predictive Modeling as a Next Step

Use the cleaned dataset to build and evaluate machine-learning classifiers (e.g., logistic regression, random forest, or gradient boosting) to quantify how well the identified features distinguish AI from human text.

9. TECHNICAL APPENDIX

A. Environment & Tools

Programming Language: Python 3.13.5

Libraries: pandas (data cleaning & analysis), numpy (numeric operations), Power BI (visualization/dashboard).

Platform: Jupyter Notebook for Python scripting; MacOS

B. Data Preparation Steps

Initial Inspection

df.shape, df.info(), and df.describe() to confirm column types and ranges.

Data Cleaning

Null handling: Imputed missing sentiment_score with mean values where appropriate.

Duplicates: df.drop_duplicates() to remove redundant rows.

Recalculated avg_word_length and avg_sentence_length to correct inconsistencies.

Validation

Cross-checked derived metrics (e.g., word_count / sentence_count ≈ avg_sentence_length).

C. Feature Engineering

- Derived additional metrics such as burstiness (sentence-length variation) and recalculated readability scores using Flesch and Gunning Fog formulas.
- Encoded categorical variables (e.g., content_type) as needed for downstream modeling.

D. Exploratory Data Analysis

- Descriptive Statistics: Grouped data by label (AI vs. human) and content_type using groupby.
- Statistical Tests: Ran two-sample t-tests on key features to check for significant differences.

E. Visualization Workflow

- Exported cleaned dataset (to_csv) for Power BI.
- Built interactive dashboards with slicers for content_type, label, and key metrics (word count, readability, burstiness, etc.).

F. Reproducibility

All Python scripts and the Power BI .pbix file are version-controlled in a private GitHub repository (Link to the repository).