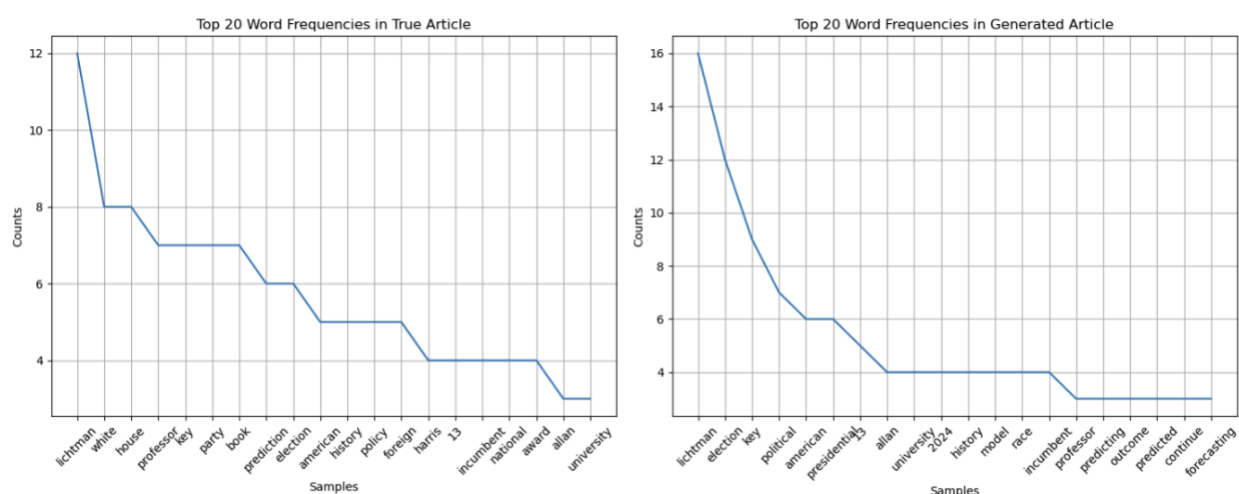Te-Jou Hou(Carol)
AU ID:5545955

The true article is entitled, "Harris or Trump? Allan Lichtman Releases His Prediction Professor Allan Lichtman's 13 Keys have been predicting presidential election winners for 40 years" and was written by Patty Housman. It was last updated on September 5, 2024.

The generated article is entitled "Allan Lichtman, Renowned American University Professor, Continues to Correctly Predict U.S. Presidential Elections."

1. Plot word frequency distributions

The true article mentions White House 8 times whereas the Generated article doesn't mention it or at least mentions it fewer than twice. The Generated text mentions political 7 times. The true text doesn't.

2. Run a grammatical analysis using Part-of-Speech Tagging

```
True Article POS Counts: Counter({'NN': 150, 'JJ': 98, 'CD': 25, 'VBD': 18, 'VBP': 11, 'RB': 9, 'VBG': 7, 'VBN': 5, 'VB': 5, 'IN': 3, 'NNS': 2, 'VBZ': 2, 'JJS': 2, 'JJR': 1, 'MD': 1, 'DT': 1})
Generated Article POS Counts: Counter({'NN': 169, 'JJ': 88, 'CD': 19, 'VBD': 16, 'VBG': 16, 'RB': 13, 'VBZ': 8, 'IN': 6, 'VBN': 6, 'VBP': 6, 'RBR': 2, 'NNS': 2, 'JJS': 2, 'DT': 1, 'RBS': 1, 'JJR': 1})
```

The distribution of verbs in the "True" text is more balanced, with a variety of tenses and forms. In contrast, the "Generated" text shows a stronger preference for the use of present participles (VBG) (True: 7, Generated: 16). The "True" text also uses numerals (CD) more frequently (True: 25, Generated: 19), suggesting a more data-driven or quantitative style. Additionally, prepositions (IN) are more common in the "Generated" text (True: 3, Generated: 6), which may indicate a tendency toward more complex sentence structures in the generated content.

3. Run a Named entity analysis

```
True Article Named Entities: 44
Generated Article Named Entities: 31
```

The "True" article contains 44 named entities, while the "Generated" article has 31. The lower number of named entities in the generated text may reflect a reduced focus on specific, real-world references, which could be indicative of the model's tendency to generate more general content, lacking in detailed references to actual people, places, or organizations.