Journal of Translation and Language Studies
https://jtls.sabapub.com

*2024 Volume 5, Issue 3: 65–83*
*DOI:* https://doi.org/10.48185/jtls.v5i3.1343

# Investigating the Linguistic Fingerprint of GPT-4o in Arabic-to-English Translation Using Stylometry

**Maysaa Banat** [1*]

[1]Languages & Liberal Arts Unit, College of Arts and Sciences, Rafik Hariri University, Mechref, Chouf 2010, Lebanon

Received: 21.08.2024 • Accepted: 28.09.2024 • Published: 30.09.2024 • Final Version: 02.12.2024

**Abstract:** This study explores the linguistic and stylistic characteristics of machine-generated texts, focusing on the output of GPT-4o. Using various natural language processing (NLP) techniques, including word frequency and stopword count analysis, readability and sentence structure metrics, lexical diversity measures, syntactic frequency analysis, and named entity recognition (NER), the research aims to uncover the stylometric fingerprints present in machine-generated content. The results reveal that GPT-4ogenerated texts exhibit moderate lexical diversity and syntactic complexity, with certain chapters reflecting higher readability and more varied sentence structures, while others lean toward simpler linguistic patterns. The findings also highlight thematic variation across chapters, as observed in the distribution of named entities, which contributes to understanding the model's handling of different contextual content.

The research suggests that while GPT-4o maintains a consistent style in its generated text, there are distinguishable characteristics that may serve as indicators of machine authorship. This provides valuable insights for stylometric analysis, authorship attribution, and the identification of machine-generated texts in various contexts. Future research could extend this work by exploring deeper stylometric features, conducting cross-model comparisons, and developing advanced authorship detection algorithms tailored for AI-generated content. Moreover, the ethical implications of stylometric analysis in the context of AI-generated texts warrant further investigation, particularly as machine-generated content becomes increasingly prevalent across different domains.

**Keywords:** Stylometric analysis, Machine-generated text, Natural Language Processing (NLP), GPT-4o, Authorship attribution

## 1. Introduction

Stylometry is the study of writing style and linguistic patterns within texts, used to analyze and quantify various aspects of an author's writing style. This field of study involves applying computational and statistical methods to large bodies of text to identify unique features, patterns, and characteristics that can differentiate one author's writing from another. Stylometry has various applications, with authorship attribution being one of the most prominent (Neal et al., 2017).

Authorship attribution is the process of determining the likely author of a text, especially in cases where the authorship is uncertain or disputed. Stylometry plays a key role in this domain by examining factors such as word choice, sentence structure, vocabulary usage, punctuation preferences, and other linguistic features. By comparing these features across different texts, stylometric techniques can provide insights into the potential authorship of a given text. It's

* Banatms@rhu.edu.lb

important to note that while stylometry can provide strong indicators, it cannot definitively prove authorship (Ramnial et al., 2016).

### 1.1. Applications of Stylometry

Stylometry has a wide array of applications, including (Neal et al., 2017):

1. **Literary Studies**: Stylometry is widely used in literary analysis to uncover hidden patterns in works of literature, identify commonalities between different texts, and even identify potential influences on authors' styles.
2. **Authorship Verification**: Beyond attribution, stylometry can help verify the authenticity of texts attributed to well-known authors. This is particularly useful when there are doubts about the true authorship of historical documents or newly discovered works.
3. **Plagiarism Detection**: Stylometry can be used to detect instances of plagiarism by comparing a text to a large corpus of existing texts. If there are significant similarities in writing style, it might indicate that the text has been copied from another source.
4. **Forensic Linguistics**: Stylometry is employed in legal and forensic contexts to determine authorship of anonymous threatening letters, ransom notes, or other texts that may be part of criminal investigations.
5. **Historical Research**: Stylometric analysis can assist in understanding the evolution of an author's style over time and in identifying anonymous or pseudonymous authors in historical documents.
6. **Psycholinguistics**: Stylometry can provide insights into an author's psychological state, personality traits, or cognitive processes based on linguistic patterns in their writing.
7. **Social Media Analysis**: Stylometry techniques can be applied to analyze and differentiate between multiple users of social media platforms, helping in tasks like identifying potential bots or analyzing user behavior.

Stylometry relies heavily on computational methods and statistical models. These include approaches like n-gram analysis (examining sequences of n words), lexical frequency analysis (counting word occurrences), syntactic analysis (examining sentence structure), and machine learning techniques for pattern recognition. Overall, stylometry is a powerful tool with a wide range of applications, particularly in the field of authorship attribution, where it aids in unraveling the mysteries of authorship and textual origins (Delcourt, 2002).

With the rise of natural language processing models, such as GPT-4o, the boundaries between human and machine-generated text have become increasingly blurred. GPT-4o, developed by OpenAI, is one of the most advanced language models to date, capable of performing a variety of tasks, including text generation, summarization, and translation. Its ability to translate documents from one language to another has opened new avenues for research, particularly in the study of linguistic patterns in machine-generated translations.

This paper seeks to investigate whether GPT-4o exhibits a unique linguistic fingerprint when translating documents from Arabic to English. Specifically, we aim to determine whether there are distinctive linguistic features in GPT-4o's translations that differentiate them from human translations. This question is of particular importance for fields such as forensic linguistics and machine-authorship detection, where identifying the origin of a text is crucial.

Stylometry relies heavily on computational methods and statistical models. These include approaches like n-gram analysis (examining sequences of n words), lexical frequency analysis (counting word occurrences), syntactic analysis (examining sentence structure), and machine learning techniques for pattern recognition. Overall, stylometry is a powerful tool with a wide range of applications, particularly in the field of authorship attribution, where it aids in unraveling the mysteries of authorship and textual origins (Delcourt, 2002).

With the rise of natural language processing models, such as GPT-4o, the boundaries between human and machine-generated text have become increasingly blurred. GPT-4o, developed by OpenAI,

is one of the most advanced language models to date, capable of performing a variety of tasks, including text generation, summarization, and translation. Its ability to translate documents from one language to another has opened new avenues for research, particularly in the study of linguistic patterns in machine-generated translations.

This paper seeks to investigate whether GPT-4o exhibits a unique linguistic fingerprint when translating documents from Arabic to English. Specifically, we aim to determine whether there are distinctive linguistic features in GPT-4o's translations that differentiate them from human translations. This question is of particular importance for fields such as forensic linguistics and machine-authorship detection, where identifying the origin of a text is crucial.

The research presented here is grounded in stylometry, drawing on various linguistic features such as vocabulary diversity, sentence structure, punctuation usage, and readability measures. By comparing GPT-4o's translations with those done by human translators, we aim to uncover whether there are quantifiable differences in style that can be used to attribute authorship or origin to machine-generated texts.

## 2. Related Work

In the age of artificial intelligence and natural language processing, the study of stylometry, authorship attribution, and the translation capabilities of advanced language models like GPT-4o has gained increasing significance. These areas of research intersect and contribute to our understanding of linguistic analysis, text generation, and the unique characteristics of machine-generated text.

Stylometry, the quantitative analysis of writing style, has a rich history dating back to the early 19th century when Edgar Allan Poe first proposed the idea of identifying authors by their unique linguistic traits. Today, stylometry plays a pivotal role in authorship attribution, forensic linguistics, and plagiarism detection. Researchers employ statistical methods and machine learning algorithms to uncover linguistic features that distinguish one writer from another.

### 2.1. Stylometry and Authorship Attribution

Stylometry has evolved from its literary origins into a field with diverse applications. Authorship attribution, a subdomain of stylometry, focuses on identifying the author of a text based on linguistic patterns. This practice is particularly useful in areas such as criminal investigations, plagiarism detection, and literary studies. Key methodologies in authorship attribution include the analysis of various linguistic features, including vocabulary, syntax, and punctuation. Machine learning models have become indispensable tools for automating this process, enabling the analysis of vast datasets and complex linguistic traits. Research in this field has demonstrated the effectiveness of stylometric techniques in accurately attributing authorship (Ramnial et al., 2016).

For effective author attribution and the detection of potential plagiarism suspects, it is crucial to understand the stylometric features employed in unbiased authorship identification. In Abbasi and Chen (2008), researchers employed a writeprints-based approach to determine the author of a given document. Their dataset consisted of online texts, including Enron emails, eBay comments, Java forum discussions, and cyber-watch chats, with experiments involving 25, 50, and 100 authors. They utilized various features such as content words, Part of Speech (PoS) tagging, word length, vocabulary richness, and sentence length for prediction. Their most accurate prediction reached 94% using the sliding window algorithm. The study also compared these results with several algorithms, including Support Vector Machine (SVM), Ensemble SVM, PCA, and Karhunen-Loeve (KL) transforms.

In a prior study by the same authors in 2006, Pavelec et al., (2008), the researchers examined 300 messages per forum in both English and Arabic, employing lexical (character-based and word-based features), syntactic (function words and PoS), structural (paragraphs and greetings), and content-based features (content words). They reported that writeprints outperformed SVM when there were at least 5 instances of an author in the training set but failed when there was only a single instance (document) of an author.

Pavelec et al. (2008) utilized a corpus of 150 Portuguese news articles with 10 authors contributing 15 texts each. They focused solely on conjunctions and employed SVM for author

prediction, achieving an accuracy of 78%. Stan´czyk and Cyran (2010) used 168 novels from two well-known Polish writers, Henryk Sienkiewicz and Boleslaw Prus, and implemented a neural network (ANN) based on function words and punctuation for author recognition. They achieved an accuracy of 95.8% when using both features together.

Furthermore, Iqbal et al. (2023) leveraged stylometric features including word length, sentence length, punctuation, vocabulary richness, function words, structural-based, and content based features to predict the author of a given document. Their analysis was conducted on the Enron Email dataset with varying numbers of emails per author (ranging from 10 to 100). They achieved an accuracy of 90% when dealing with 5 authors and utilizing the k-means clustering algorithm.

### 2.2. Translation Capabilities of LLMs

The translation capabilities of LLMs, particularly GPT-4, have also been extensively explored. GPT-4 has been shown to excel in post-editing machine translations, enhancing translation quality and correcting errors. However, its translations also exhibit stylistic tendencies that differ from human translators, creating a distinct linguistic fingerprint even in high-quality translations (Raunak, 2023). The ability of LLMs to evaluate translation quality has also been studied, with models like GPT-4 and LLaMA acting as evaluators of translation output. These models tend to show a preference for specific syntactic structures, especially when translating between complex language pairs such as English-German and Chinese-English, thus reinforcing their linguistic signatures in translation tasks (Kocmi & Federmann, 2023).

In another study, Jio et al. (2023) investigated the use of ChatGPT for machine translation, encompassing aspects like translation prompts, multilingual translation, and translation robustness. Through assessments conducted on various benchmark test sets, they observe that ChatGPT demonstrates competitive performance when compared to commercial translation tools (such as Google Translate) for well-resourced European languages. However, it exhibits noticeable disparities when dealing with languages that have limited resources or are significantly different. To address these challenges with distant languages, they explore a novel approach known as "pivot prompting," where ChatGPT is instructed to first translate the source sentence into a high-resource pivot language before translating it into the target language. This strategy notably enhances translation performance. Regarding translation robustness, ChatGPT's performance is not as strong as commercial systems when dealing with biomedical abstracts or Reddit comments. Still, it yields favorable outcomes in the context of spoken language. Furthermore, with the introduction of the GPT-4 engine, ChatGPT's translation capabilities have seen substantial improvements, now making it a viable competitor to commercial translation products, even for distant languages.

The investigation into GPT-3.5's capacity for translating specialized religious texts has yielded promising outcomes. Nonetheless, it is imperative to acknowledge that additional research is needed to gain a comprehensive understanding of the model's limitations within this domain. While GPT-3.5 has displayed a reasonable degree of precision and fluency in rendering religious texts, there is a fundamental need for enhancements in its performance and precision. Future research endeavors might center on refining the training data employed to instruct machine translation models, thereby enabling them to handle domain-specific lexicon and terminology more effectively. Additionally, new evaluation metrics should be devised to accurately assess translation quality. Furthermore, concentrated efforts can be directed at augmenting the model's proficiency in dealing with intricate sentence structures and discourse elements frequently found in religious texts. By addressing these challenges, GPT-3.5 has the potential to evolve into a valuable instrument for translating specialized religious content (Banat & Abu Adla, 2023).

Multilingual fine-tuning has been employed to improve the translation capabilities of models like XGLM and GPT-4. Research shows that, even when fine-tuned on multiple languages, these models still exhibit unique stylistic patterns depending on the language pairs involved. For instance, GPT-4's translations between low-resource languages such as Arabic-Swahili tend to display a more pronounced linguistic fingerprint compared to more common language pairs like English-German (Zhu et al., 2023). Moreover, while GPT-4 consistently ranks highly on translation quality metrics, it still shows a tendency to produce stylistically distinct translations, characterized by syntactic

complexity and specific lexical choices, making its machine-generated translations distinguishable from human ones (Kocmi & Federmann, 2023).

### 2.3. Comparison between GPT-4o and Human Translation

The ongoing debate surrounding machine translation (MT) models like GPT-4o and human translation has sparked numerous comparisons of their abilities in terms of accuracy, fluency, and adaptability to context. Recent studies indicate that while GPT-4o demonstrates significant advancements in machine translation, it still lags behind human translators in several key areas, though it performs comparably to junior human translators in terms of total errors made.

GPT-4o often struggles with more complex and nuanced translations, particularly when moving between resource-poor languages or handling context-dependent idiomatic expressions. Machine translations tend to be more literal, reflecting the model's focus on word-for-word translation rather than grasping the deeper context that human translators typically capture. This literalness can lead to translations that are technically correct but lack the intended meaning or tone (Yan et al., 2024).

Human translators excel in understanding the broader context of a text, making adjustments based on cultural or contextual clues. Studies show that while GPT-4o can handle straightforward technical translations relatively well, it often fails to interpret metaphors, idioms, or stylistic variations, which are critical for producing natural, human-like translations. Humans, conversely, can "overthink" certain translations by injecting more context than is present in the original text, sometimes adding unnecessary background information (Yan et al., 2024).

Human translations are often more flexible in their structure, as humans can adapt sentence length, syntactic patterns, and word choice based on cultural and linguistic expectations. GPT-4o tends to follow a more rigid pattern, which can lead to awkward phrasing, particularly in longer sentences. Studies suggest that GPT-4o's translations, while grammatically sound, often lack the stylistic variation and fluidity that human translations naturally exhibit (Son & Kim, 2023).

When considering less commonly spoken languages or those with fewer available training datasets, human translators outperform GPT-4o by a wide margin. For languages with more resources, such as English-German or French-English, GPT-4o shows competitive results, occasionally surpassing human translations in technical contexts. However, for resource-poor languages, the performance gap widens significantly, with GPT-4o displaying higher error rates and less reliable fluency (Son & Kim, 2023).

In summary, GPT-4o excels in certain structured, domain-specific tasks but remains inferior to human translators in terms of capturing nuance, context, and cultural subtleties. As MT models continue to evolve, we may see improvements in these areas, but for now, human translators remain essential for high-quality, contextually accurate translations.

Future work should focus on enhancing the contextual understanding of GPT-4o by incorporating diverse and underrepresented languages into training datasets. Additionally, hybrid models, where human and machine translators work in tandem, may help address the weaknesses of both systems.

### 2.4. Linguistic Fingerprints in LLMs

Linguistic fingerprints in large language models (LLMs) have emerged as a critical area of research, especially in the context of authorship attribution and stylistic analysis. Various studies have demonstrated that LLMs like GPT-4, LLaMA, and GPT-NeoX exhibit unique linguistic patterns that distinguish them from both human authors and each other. One approach to detecting these fingerprints involves using stylometric watermarks embedded within the text generation process, which manipulate token probabilities to create detectable patterns (Niess & Kern, 2024). These stylometric markers serve as a form of authorship verification, making it possible to trace the origin of machine-generated text. Neural authorship attribution studies have also found that lexical features such as word length, vocabulary richness, and syntactic complexity play a significant role in differentiating between texts produced by various LLMs. For example, texts generated by GPT-4 and LLaMA show distinct stylistic differences, even though both models were trained on similar

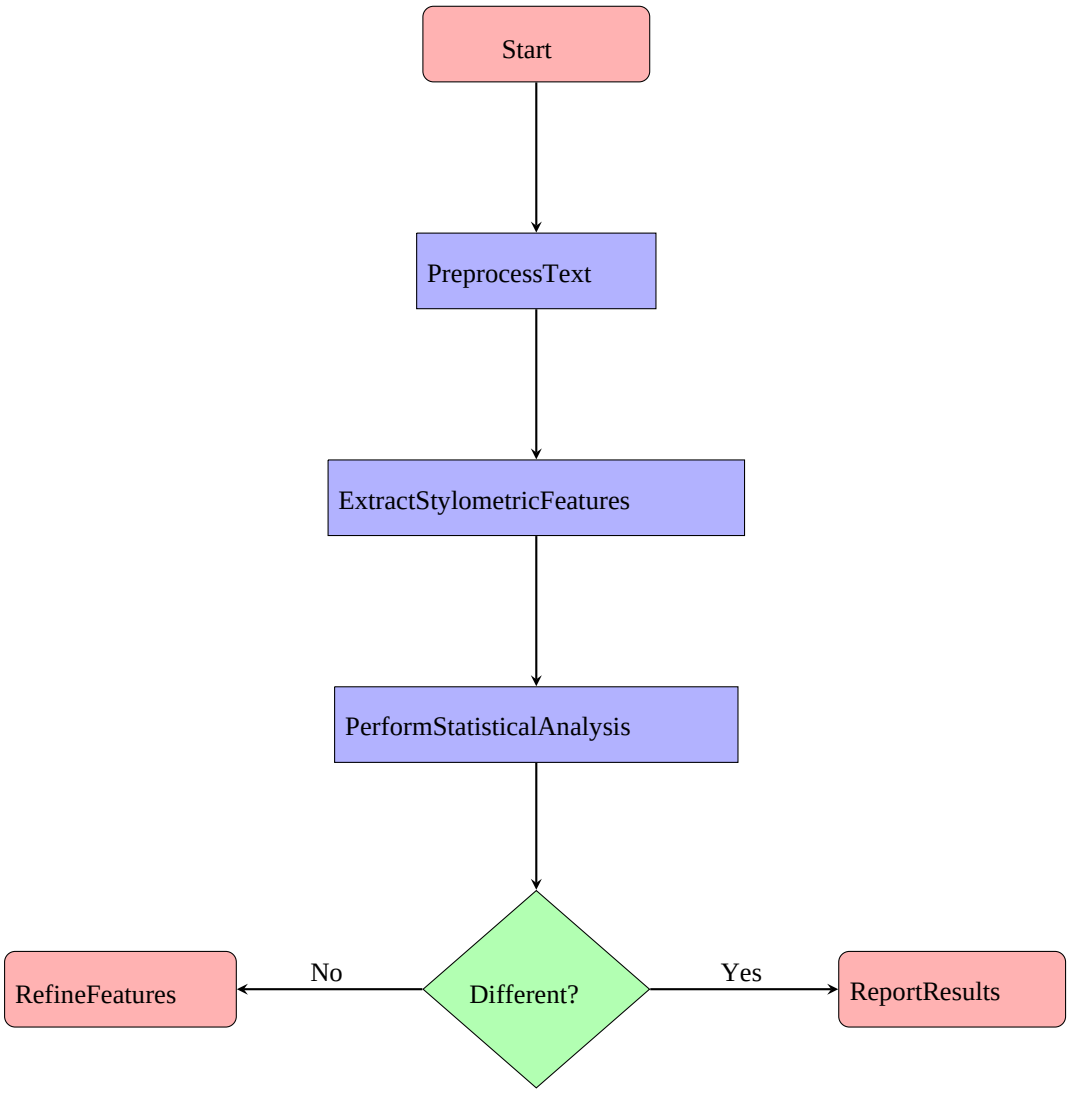
datasets (Kumarage & Liu, 2023). This is further supported by research that examines how fine-tuning LLMs for specific tasks alters their linguistic signatures, contributing to a unique "fingerprint" for each model version (Diwan, 2021).

Another study commenced by conducting a comparison of Japanese stylometric characteristics between texts generated by GPT-3.5 and GPT-4 and those authored by humans (Zaitsu & Jin, 2023). Their approach involved multi-dimensional scaling (MDS) to examine the distribution patterns of 216 texts categorized into three groups: 72 academic papers written by 36 individual authors, 72 texts generated by GPT-3.5, and 72 texts generated by GPT-4. The researchers focused on several key stylometric features, namely: (1) bigrams of parts-of-speech, (2) bigrams of postpositional particle words, (3) comma placement, and (4) the frequency of function words. The MDS analysis revealed distinct distribution patterns for each stylometric feature among texts generated by GPT (-3.5 and -4) and those produced by human authors. Notably, despite GPT-4's enhanced computational capabilities due to a larger number of parameters, the distributions of both GPT versions exhibited overlap. These findings suggest that even as AI models evolve with increased parameter counts in the future, text generated by GPT may not closely resemble human-authored content in terms of stylometric characteristics.

Instructional fingerprinting is another method that highlights how the style of LLM outputs can vary based on the input prompts given to the model. This research shows that even subtle changes in instructions can lead to significant stylistic shifts in the generated content, thus contributing to a distinct linguistic signature (Xu et al., 2024). Additionally, studies comparing GPT-3 and human writers reveal that, despite GPT-3's ability to closely mimic human styles, its generated text still retains identifiable linguistic markers, making it possible to distinguish between machine-generated and human-authored content (Zaistsu & Jin, 2023)

## 3. Methodology

The proposed method consists of a systematic approach to analyzing the linguistic fingerprint of GPT-4o translations using various stylometric features. The process involves extracting multiple linguistic features from the translated documents and conducting statistical analyses to identify significant differences. The overall methodology is summarized in the flowchart below.




### 3.1. Dataset Description

The dataset used in this study serves as a valuable resource for evaluating the effectiveness of NLP models in translating specialized religious texts from Arabic to English. The dataset comprises a religious book spanning 239 pages, containing 30 chapters in total, along with an introductory chapter and a concluding section. Each chapter consists of approximately 7 pages, with around 5 paragraphs per page, resulting in an average of 1,500 words per chapter. The religious content in the book includes Hadith (sayings of the Prophet Muhammad) and Quranic verses, along with verification and sourcing information for the Hadith (Farshoukh, 2018).

For this study, we focused on 10 chapters from the book. These chapters were translated by both human experts and the GPT-4o model, providing a basis for comparison.

#### 3.1.1. Human-Translated Chapters

The human-translated chapters are approximately 9 pages in length, with around 5 paragraphs per page, averaging around 1,800 words per chapter. These translations include Hadith, Quranic verses, and their corresponding Hadith verification and sourcing.

### 3.1.2. GPT-4o Translated Chapters

The GPT-4o translated chapters are slightly shorter, at around 8 pages per chapter, with 5 paragraphs per page, resulting in an average word count of approximately 1,700 words per chapter. These chapters also include Hadith, Quranic verses, and Hadith verification and sourcing, allowing for a thorough comparison of the translation quality between human translators and GPT-4o.

This dataset offers an ideal platform for analyzing how GPT-4o handles the nuances of translating religious texts, particularly regarding the preservation of meaning, context, and cultural sensitivity.

## 3.2. GPT-4o: Architecture, Training, and Capabilities

GPT-4o is a large-scale language model designed to excel at natural language understanding and generation tasks, building upon the advances of its predecessor, GPT-4. The architecture of GPT-4o, like other GPT models, is based on the Transformer architecture, which uses self-attention mechanisms to process and generate sequences of text. This allows the model to handle a wide range of linguistic patterns, from simple sentence structures to more complex, context-dependent expressions (Islam & Moushi, 2024).

The model size of GPT-4o is vast, with billions of parameters, making it highly capable of capturing intricate patterns in language. Each parameter represents a weight in the neural network that has been fine-tuned based on the training data. The large number of parameters allows GPT-4o to handle complex linguistic tasks, including translation, summarization, and question-answering, with higher accuracy compared to earlier models (Islam & Moushi, 2024).

GPT-4o's training data consists of a diverse range of text sources, including books, articles, and websites, making it adept at generalizing across multiple domains. However, its performance is influenced by the amount of training data available for specific languages and tasks. The model benefits from large-scale, unsupervised training, where it learns to predict the next word in a sentence based on the previous words, thereby acquiring a robust understanding of syntax, semantics, and world knowledge (Islam & Moushi, 2024).

Despite its strengths, GPT-4o, like all large language models, faces challenges, particularly in handling less-represented languages and generating nuanced translations that require deep cultural understanding. The vast computational resources required to train and fine-tune models of this size are another key consideration in its deployment (Islam & Moushi, 2024).

In conclusion, GPT-4o is a powerful tool for natural language tasks, leveraging its large model size, Transformer-based architecture, and extensive training data to deliver impressive results. However, it is not without limitations, particularly in areas that require human-like reasoning or cultural insight.

## 3.3. Preprocessing and Feature Extraction

Before performing stylometric analysis, several text preprocessing steps were applied to the translated documents to standardize the data and ensure consistency. The preprocessing phase included tokenization, where the text was split into individual words, and the removal of punctuation, which could interfere with feature extraction. We also converted the text to lowercase to avoid case sensitivity issues and removed any unnecessary whitespace or special characters.

Following preprocessing, we extracted key linguistic features for stylometric analysis. These features included vocabulary richness, measured using metrics like Type-Token Ratio (TTR) and Herdan’s C, to assess the diversity of words used in the translations. Sentence length was another critical feature, which helped in evaluating the complexity of the text, along with readability indices like the Flesch-Kincaid Grade Level and the Gunning Fog Index. We also analyzed punctuation usage, particularly focusing on the frequency of commas, periods, and other marks, as punctuation often reflects stylistic tendencies. Lastly, syntactic features, such as part-of-speech (PoS) tagging and named entity recognition (NER), were extracted to provide deeper insights into the grammatical structures and named entities within the text, further distinguishing the stylistic patterns of human versus GPT-4o translations.

### 3.4. Metrics Definition

In this study, we computed several linguistic metrics to capture different aspects of the text's style, including word usage, sentence structure, and syntactic features. These metrics help quantify differences between human-translated texts and GPT-4o translations. The following subsections define the key metrics computed.

#### 3.3.1. Word Usage

**Most Frequent Words**: This metric identifies the most common words used in the translated text. We calculated the frequency of the top 10 most frequently occurring words, denoted as:

$$f(w_i) = \frac{\text{count of word } w_i}{\text{total number of words in the document}}$$

Where $w_i$ represents the $i$-th word in the document.

**Stopword Usage**: Stopwords (common words like "and," "the," "is," etc.) were counted to assess their frequency. The list of stopwords was derived from the Natural Language Toolkit (NLTK) library. The stopword usage ratio is given by:

$$\text{Stopword Ratio} = \frac{\text{Number of stopwords}}{\text{Total number of words}}$$

This ratio helps identify the extent to which common words contribute to the overall word count.

#### 3.3.2. Sentence Structure

To evaluate sentence complexity, we computed several readability metrics, as defined below:
**Average Sentence Length (ASL)**: The average number of words per sentence is calculated as:

$$\text{ASL} = \frac{\text{Total number of words}}{\text{Total number of sentences}}$$

**Flesch-Kincaid Grade Level (FKGL)**: This readability metric estimates the U.S. grade level required to comprehend the text. It is calculated using the formula:

$$\text{FKGL} = 0.39 \times \left(\frac{\text{Total words}}{\text{Total sentences}}\right) + 11.8 \times \left(\frac{\text{Total syllables}}{\text{Total words}}\right) - 15.59$$

**Gunning Fog Index**: This index estimates the number of years of formal education required to understand the text. It uses the following formula:

$$\text{GFI} = 0.4 \times \left(\frac{\text{Total words}}{\text{Total sentences}} + \frac{\text{Complex words}}{\text{Total words}} \times 100\right)$$

Where "complex words" are defined as words with three or more syllables.

**Automated Readability Index (ARI)**: ARI calculates readability based on word and character counts, using the formula:

$$\text{ARI} = 4.71 \times \left(\frac{\text{Total characters}}{\text{Total words}}\right) + 0.5 \times \left(\frac{\text{Total words}}{\text{Total sentences}}\right) - 21.43$$

### 3.3.3. Lexical Diversity

Lexical diversity measures the richness of vocabulary in the text. We calculated two metrics:
**Type-Token Ratio (TTR)**: The ratio of unique words (types) to the total number of words (tokens) in the text is calculated as:

$$\text{TTR} = \frac{\text{Number of unique words (types)}}{\text{Total number of words (tokens)}}$$

**Herdan's C**: This is another measure of lexical diversity, less sensitive to text length, given by:

$$\text{Herdan's C} = \frac{\log(\text{Number of types})}{\log(\text{Number of tokens})}$$

Herdan's C provides a more stable measure of vocabulary richness across texts of varying lengths.

### 3.3.4. Syntactic Features

To analyze syntactic features, we used Part of Speech (PoS) tagging to compute the frequency of different syntactic tags in the text. The syntactic features analyzed are essential in determining the structural complexity of the translations. The following list outlines the tags used in this analysis:

- **det**: Determiner (e.g., "the," "a")
- **nsubj**: Nominal subject (e.g., the subject of the sentence)
- **ccomp**: Clausal complement (e.g., clauses that complement a verb)
- **neg**: Negation (e.g., "not," "no")
- **prep**: Preposition (e.g., "in," "on")
- **amod**: Adjectival modifier (e.g., adjectives modifying nouns)
- **pobj**: Object of preposition (e.g., "in the mosque" — "mosque" is the pobj)
- **punct**: Punctuation
- **poss**: Possession modifier (e.g., "Muhammad's word" — "Muhammad's" is poss)
- **nsubjpass**: Passive nominal subject (e.g., the subject in a passive construction)
- **aux**: Auxiliary verb (e.g., "is," "has")
- **auxpass**: Auxiliary verb in passive voice (e.g., "was" in "was written")
- **ROOT**: Root of the sentence (main verb or predicate)
- **cc**: Coordinating conjunction (e.g., "and," "but")
- **conj**: Conjunct (e.g., elements connected by conjunctions)
- **attr**: Attribute (e.g., a characteristic or property of the subject)
- **dobj**: Direct object (e.g., the object acted on by the verb)
- **advcl**: Adverbial clause modifier (e.g., a clause that modifies a verb)
- **compound**: Compound (e.g., multi-word names or expressions)
- **appos**: Appositional modifier (e.g., renaming a noun phrase)
- **advmod**: Adverbial modifier (e.g., an adverb modifying a verb or adjective)
- **mark**: Marker (e.g., introduces a subordinate clause)
- **xcomp**: Open clausal complement (a clause without its own subject)
- **relcl**: Relative clause modifier (e.g., a clause modifying a noun)
- **acl**: Adjectival clause modifier (e.g., a clause modifying a noun)

These syntactic features were used to analyze the complexity and structure of the translated chapters, allowing us to compare the outputs across different chapters. The frequency of these syntactic tags

provides insight into the sentence structure and grammatical patterns employed by GPT-4o in the translations.

#### 3.3.5. Named Entity Recognition (NER)

Named Entity Recognition (NER) is a process in NLP that identifies and categorizes named entities (such as people, locations, organizations, dates, and other specific entities) in a text. NER plays a crucial role in extracting structured information from unstructured text by detecting and classifying key pieces of information.
For this analysis, we extracted the following types of entities:

- **PERSON**: Refers to individual people or characters.
- **GPE**: Geo-political entities, including countries, cities, and states.
- **ORG**: Organizations, such as companies, government bodies, or institutions.
- **DATE**: Temporal expressions, including specific dates or periods.
- **CARDINAL**: Numeric references, such as quantities or counts.
- **WORK OF ART**: Titles of books, movies, paintings, or other creative works.
- **LOC**: Specific locations (e.g., geographical areas that are not geopolitical).
- **ORDINAL**: First, second, etc., indicating order or ranking.
- **NORP**: Nationalities, religious groups, or political groups.
- **FAC**: Physical facilities, including buildings, airports, or highways.
- **TIME**: Specific times, such as hours or periods during the day.
- **EVENT**: Refers to notable events, such as festivals or wars.

In this study, NER was employed to assess the frequency of these entities across different chapters, allowing us to capture a detailed representation of the text's content and thematic elements. By analyzing the named entities, we can further understand the focus of each chapter, whether it is centered around individuals, locations, organizations, or temporal information.

## 4. Results and Discussion

This section presents the outcomes of the analyses conducted on the text, encompassing word frequency and stopword count, readability and sentence structure, lexical diversity, syntactic frequency, and NER. These analyses provide a comprehensive understanding of the linguistic and syntactic patterns within the text, offering insights into both surface level features, such as word usage, and deeper structures, such as sentence complexity and syntactic constructs. The results shed light on the stylistic tendencies of the text, which could be linked to the distinct linguistic fingerprint associated with different chapters or sections, thereby contributing to stylometric analysis. The following subsections discuss the key findings and their implications for authorship attribution and textual analysis.

### 4.1. Word Frequency and Stopword Count Analysis

In machine-generated texts, especially those created using models such as GPT-4o, certain patterns of word usage, including the distribution of function words (i.e., stopwords) and content words, can form a "linguistic fingerprint" that helps differentiate between human and machine authorship.

In Figure 1, the word frequencies and stopword counts from five different chapters (Chapter 11, 12, 14, 17, and 18) are visualized. The graph shows a relatively consistent pattern in the proportion of stopwords to total words across the chapters, with some chapters having slightly higher proportions of stopwords.

As seen in the plot, Chapter 14 shows the highest total word count, with a noticeable dominance of stopwords over content words. This aligns with patterns often seen in formal or verbose human writing, where sentence connectors and filler words are used more frequently to maintain a natural flow. Chapters 11 and 12 show slightly lower word frequencies but maintain a similar proportion between stopwords and content words. This balance suggests a more machine-like generation, potentially indicative of algorithmic text production where filler words are less frequent, and the focus is placed on delivering information-rich content. On the other hand, Chapter 18 demonstrates a moderate word count and stopword distribution. The ratio between stopwords and content words suggests a balance that could either be characteristic of less formal human writing or a machine-generated text that mimics a conversational tone.

The differences in stopword usage and word frequency across the chapters provide a basis for stylometric analysis. The relatively consistent proportion of stopwords across the chapters may suggest a shared authorship or generation style. However, the slight variations in frequency and the balance between content and stopwords could indicate subtle stylistic shifts that reflect either topic variations or an evolving machine generation model.

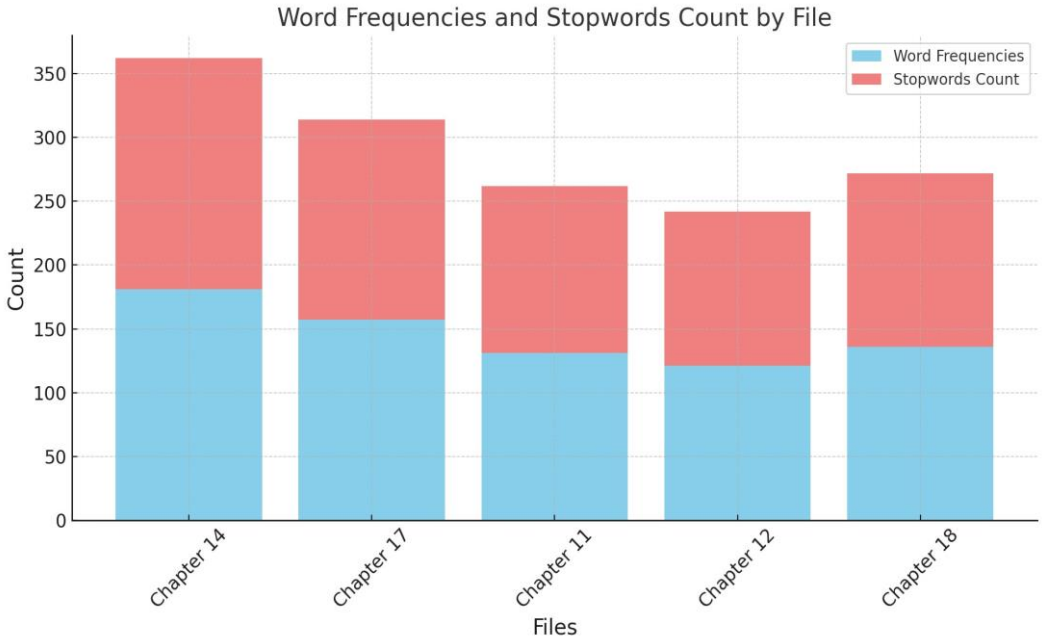



**Figure 1.** Word Frequencies and Stopwords Count by File

### 4.2. Readability and Sentence Structure Analysis

In this section, we analyze various sentence-level metrics and readability indices across the different chapters of the GPT-4o translations. These metrics provide insights into the complexity and readability of the translated text, helping us understand the stylometric fingerprint of GPT-4o.

As shown in Figure 2, the average sentence length varies across the chapters, ranging from 18.68 words per sentence (Chapter 12) to 23.29 words (Chapter 9). This variability indicates differences in the sentence complexity employed by GPT-4o in each chapter. Chapters with longer average sentence lengths, such as Chapters 9 and 18, tend to have higher readability scores.

The Flesch-Kincaid Grade Level scores reveal that most chapters fall between the 6th and 10th-grade reading levels, with Chapter 12 being the most accessible (6.9) and Chapter 9 being the most complex (10.3). Similarly, the Gunning Fog Index shows a corresponding trend, with values ranging from 8.62 (Chapter 12) to 12.07 (Chapter 9).

The Automated Readability Index (ARI) also supports these observations, with Chapters 9 and 18 showing higher ARI scores, indicating that these chapters may require a higher level of reading proficiency.

These metrics together suggest that while GPT-4o translations generally maintain a moderate level of readability, certain chapters (e.g., Chapters 9 and 18) exhibit higher complexity due to longer sentences and more sophisticated language structures.


**Figure 2.** Readability and Sentence Metrics across Chapters

### 4.3. Lexical Diversity Analysis

Lexical diversity is a measure of the richness and variety of vocabulary used in the text. In this study, we computed two key metrics: the TTR and Herdan's C. These metrics provide insights into the lexical variety present in the translations across different chapters.

A higher TTR indicates greater lexical diversity. However, TTR is sensitive to text length, which is why Herdan's C is also used, as it accounts for text length and provides a more stable measure of lexical diversity across varying text lengths.

The results for TTR and Herdan's C across different chapters are presented in the table below, ordered by chapter number:

**Table 1.** Lexical Diversity: Type-Token Ratio and Herdan's C for Different Translated Chapters, Ordered by Chapter Number

| File | Type-Token Ratio (TTR) | Herdan's C |
|---|---|---|
| Chapter 9 | 0.4007 | 8.0863 |
| Chapter 10 | 0.3844 | 7.8522 |
| Chapter 11 | 0.3615 | 7.5380 |
| Chapter 12 | 0.3811 | 7.8172 |
| Chapter 13 | 0.3437 | 7.0821 |

| | | |
|---|---|---|
| Chapter 14 | 0.3720 | 7.6732 |
| Chapter 15 | 0.3714 | 7.4825 |
| Chapter 16 | 0.3347 | 7.2079 |
| Chapter 17 | 0.3842 | 7.9587 |
| Chapter 18 | 0.3999 | 8.2242 |

As shown in Table 1, it is evident that Chapter 9 and Chapter 18 exhibit the highest lexical diversity, with TTR values of 0.4007 and 0.3999, respectively, and corresponding Herdan's C values of 8.0863 and 8.2242. This suggests that these chapters contain a wider range of vocabulary, indicating a richer lexical usage compared to other chapters.

In contrast, Chapter 16 shows the lowest TTR of 0.3347 and a Herdan's C of 7.2079, suggesting less lexical variety. These results imply that Chapter 16 may rely more on repetitive word use or simpler language structures, whereas Chapters 9 and 18 employ more diverse and complex vocabulary.

The combination of TTR and Herdan's C provides a comprehensive view of the lexical diversity in the translations, with Herdan's C serving as a more robust metric for comparing texts of different lengths.

To assess the overall lexical diversity and stylistic tendencies of GPT-4o translations, we calculated the averages of the TTR and Herdan's C across all chapters. These averages provide insights into the general stylistic patterns present in the translated documents.

The average TTR of 0.3734 indicates that while GPT-4o exhibits moderate lexical diversity, the model tends to reuse words more frequently than expected in highly creative or human-authored texts, where TTR values might be higher. This suggests that the translations, while consistent, may not employ as varied a vocabulary as human translators might in similar contexts.

The average Herdan's C of 7.6922 supports this observation, suggesting that while GPT4o is capable of introducing diverse vocabulary, it still falls within a relatively stable range of lexical complexity across different chapters. Chapters with higher Herdan's C, such as Chapter 9 and Chapter 18, reflect instances where GPT-4o employs a broader range of vocabulary, whereas chapters with lower values, like Chapter 16, show simpler linguistic structures.

Overall, these findings indicate that GPT-4o translations, while maintaining coherence and structural consistency, exhibit moderate lexical diversity and stylistic repetition. These characteristics align with the model's tendency to generate fluent yet formulaic translations, a key feature of machine-generated text. The results highlight that GPT-4o, although powerful in generating translations, still reflects a stylometric fingerprint distinct from human translators, particularly in terms of lexical variety and complexity.

### 4.4. Syntactic Frequency Analysis

The table below presents the frequency of various syntactic components across different chapters translated by ChatGPT.

**Table 2.** Syntactic Tag Frequencies Across Translated Chapters

| **Syntactic Tag** | **Ch.9** | **Ch.10** | **Ch.11** | **Ch.12** | **Ch.13** | **Ch.14** | **Ch.15** | **Ch.16** | **Ch.17** | **Ch.18** |
|---|---|---|---|---|---|---|---|---|---|---|
| det | 152 | 206 | 192 | 167 | 173 | 217 | 187 | 225 | 205 | 198 |
| nsubj | 175 | 189 | 275 | 225 | 235 | 222 | 175 | 305 | 190 | 182 |
| ccomp | 43 | 57 | 73 | 58 | 59 | 65 | 40 | 82 | 43 | 53 |
| neg | 12 | 12 | 15 | 15 | 15 | 15 | 17 | 26 | 27 | 7 |
| prep | 204 | 229 | 220 | 224 | 237 | 243 | 196 | 331 | 237 | 235 |
| amod | 34 | 35 | 50 | 39 | 34 | 39 | 27 | 42 | 68 | 47 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| pobj | 192 | 225 | 218 | 219 | 225 | 237 | 195 | 321 | 226 | 233 |
| punct | 300 | 362 | 445 | 392 | 359 | 403 | 318 | 539 | 343 | 378 |
| poss | 62 | 51 | 82 | 61 | 63 | 66 | 50 | 64 | 65 | 53 |
| nsubjpass | 13 | 20 | 20 | 10 | 24 | 15 | 12 | 26 | 22 | 14 |
| aux | 69 | 66 | 71 | 79 | 80 | 43 | 69 | 111 | 95 | 78 |
| auxpass | 15 | 20 | 22 | 15 | 23 | 15 | 13 | 31 | 21 | 13 |
| ROOT | 70 | 80 | 95 | 101 | 97 | 96 | 82 | 128 | 95 | 85 |
| cc | 97 | 101 | 116 | 90 | 84 | 127 | 85 | 159 | 134 | 98 |
| conj | 109 | 99 | 126 | 98 | 86 | 139 | 97 | 158 | 167 | 96 |
| attr | 23 | 21 | 31 | 30 | 19 | 25 | 26 | 26 | 26 | 19 |
| dobj | 108 | 104 | 128 | 105 | 119 | 130 | 102 | 177 | 130 | 130 |
| dep | 23 | 20 | 16 | 19 | 21 | 39 | 33 | 40 | 39 | 32 |
| relcl | 37 | 18 | 45 | 25 | 32 | 37 | 30 | 38 | 32 | 27 |
| advcl | 29 | 39 | 55 | 41 | 40 | 40 | 38 | 50 | 38 | 43 |
| appos | 16 | 28 | 24 | 18 | 15 | 24 | 12 | 19 | 5 | 42 |
| compound | 21 | 47 | 63 | 78 | 83 | 32 | 29 | 33 | 13 | 27 |
| npadvmod | 1 | 10 | 11 | 3 | 8 | 6 | 6 | 9 | 2 | 9 |
| expl | 6 | 3 | 6 | 4 | 1 | 4 | 4 | 2 | 1 | 1 |
| prt | 2 | 4 | 2 | 8 | 6 | 8 | 7 | 12 | 8 | 6 |
| advmod | 54 | 60 | 81 | 75 | 68 | 59 | 63 | 97 | 39 | 66 |
| preconj | 4 | 2 | 1 | 1 | 1 | 3 | 0 | 1 | 1 | 2 |
| csubj | 5 | 1 | 0 | 2 | 0 | 3 | 4 | 7 | 3 | 0 |
| csubjpass | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 0 | 0 |
| mark | 26 | 32 | 41 | 29 | 40 | 36 | 24 | 63 | 28 | 40 |
| intj | 1 | 3 | 9 | 5 | 3 | 2 | 3 | 7 | 0 | 2 |
| case | 5 | 8 | 2 | 7 | 6 | 5 | 1 | 8 | 12 | 6 |
| pcomp | 12 | 4 | 11 | 6 | 8 | 10 | 5 | 26 | 28 | 9 |
| acomp | 13 | 9 | 28 | 23 | 20 | 13 | 10 | 26 | 22 | 8 |
| acl | 5 | 4 | 6 | 5 | 3 | 3 | 6 | 9 | 6 | 6 |
| nmod | 0 | 4 | 0 | 3 | 1 | 4 | 4 | 2 | 1 | 4 |
| dative | 5 | 11 | 9 | 0 | 5 | 1 | 4 | 18 | 4 | 3 |
| oprd | 2 | 4 | 1 | 4 | 1 | 1 | 2 | 1 | 1 | 2 |
| nummod | 7 | 9 | 3 | 10 | 5 | 3 | 5 | 12 | 5 | 14 |
| agent | 3 | 3 | 7 | 3 | 3 | 2 | 1 | 3 | 4 | 2 |
| parataxis | 1 | 1 | 3 | 9 | 5 | 1 | 7 | 7 | 4 | 6 |
| xcomp | 9 | 21 | 11 | 17 | 17 | 1 | 17 | 18 | 23 | 25 |
| predet | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 |
| meta | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| quantmod | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

The syntactic analysis, in Table 2, reveals notable differences in the frequency of syntactic tags across the translated chapters. For instance, noun phrases (NN) were consistently the most frequent syntactic element across all chapters, peaking in Chapter 16 (305 occurrences) and being lowest in Chapter 9 (175 occurrences). This suggests that Chapter 16 contained a higher concentration of subject or object nouns, reflecting its likely descriptive nature.

The use of determiners (det), such as “the” or ”a,” showed relatively high frequency across chapters, with Chapter 16 again having the highest frequency (225 occurrences). This consistency indicates a stable use of definite and indefinite articles across translations, which is crucial for grammatical coherence in both human and machine-translated texts.

Conversely, lower frequencies in syntactic tags like negation (neg) and auxiliary verbs (auxpass) suggest that certain linguistic constructs, such as passive voice and negation, were used less frequently, particularly in Chapters 18 and 12. The relatively low use of negation could indicate that the content

in these chapters contained more declarative statements rather than negations or passive constructions. Additionally, punctuation (punct) varied significantly across chapters, with Chapter 16 showing a much higher frequency (539 occurrences), likely due to complex sentence structures requiring more punctuation.

These results indicate that, while there is a general consistency in core syntactic structures (e.g., nouns, determiners), there are distinct differences in the use of more specific tags, which might reflect the differing complexity and focus of the chapters. This variation also highlights how GPT-4o handles syntactic structures across different contexts, providing insights into the style and consistency of its translations.

### 4.5. NER Analysis

In this section, we analyze the distribution of named entities identified by the GPT-4o model across the different translated chapters. Understanding the distribution of these entities can help identify thematic differences in the chapters and uncover any potential biases or patterns in the translation process.

The results of the NER analysis for each chapter are visualized in Figure 3, which presents the entity counts in a stacked bar chart for ease of comparison. The chart breaks down the entity counts by chapter, with each colored section representing a different entity type.

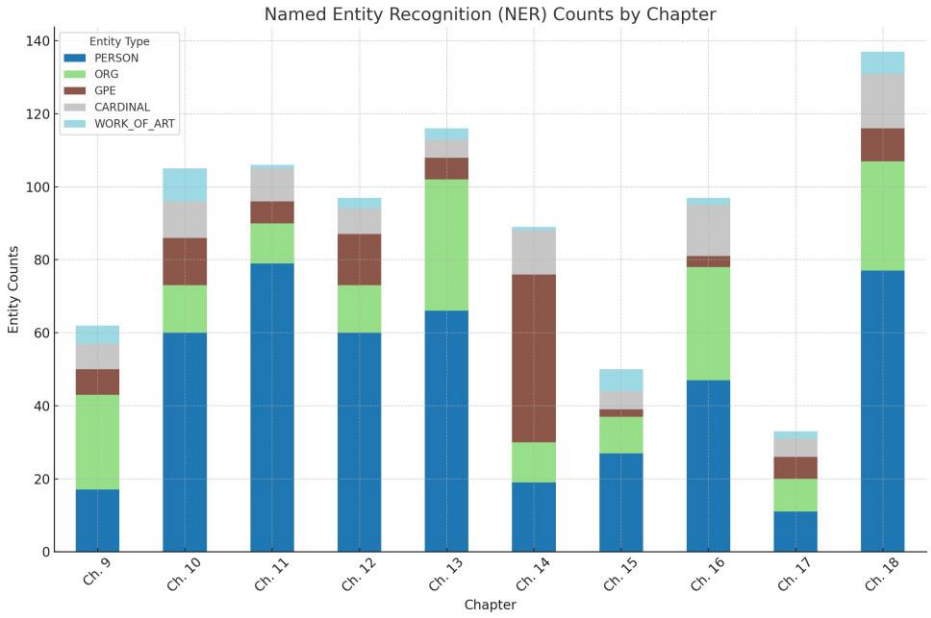



**Figure 3.** Named Entity Recognition (NER) Counts by Chapter

The analysis reveals significant variation in the distribution of entities across chapters.

PERSON entities dominate most chapters, with the highest counts observed in Chapter 11 (79 occurrences) and Chapter 18 (77 occurrences). This suggests that these chapters are heavily focused on individual people, potentially due to the nature of the content.

ORG (Organization) entities also play a significant role in chapters such as Chapter 13, which has 36 occurrences, indicating a focus on institutions or organizations. CARDINAL entities, representing numerical data, are most frequent in Chapters 16 and 18, reflecting a higher prevalence of numerical references in these sections.

Notably, Chapter 14 contains the highest number of GPE mentions (46 occurrences), indicating a strong focus on locations or geopolitical contexts. This is in stark contrast to chapters like Chapter 15, where GPE is almost negligible (only 2 occurrences).

The entity type WORK OF _ART, representing references to artworks or significant cultural texts, is less frequent overall but appears consistently across most chapters, with peaks in Chapter 10 (9 occurrences) and Chapter 18 (6 occurrences).

Overall, the distribution of entities reflects the thematic diversity across the chapters. Chapters that focus on individuals and their actions show higher counts for PERSON entities, while others that discuss organizations or geopolitical contexts are reflected by higher counts in ORG and GPE entities. This variation highlights the capacity of GPT-4o to capture and translate distinct thematic elements based on the content.

## 5. Conclusion and Future Directions

This study aimed to explore the distinct linguistic and stylistic patterns of machine-generated texts using various NLP analyses, including word frequency and stopword count, readability and sentence structure, lexical diversity, syntactic frequency, and NER. Through this analysis, we sought to uncover potential stylometric fingerprints present in texts generated by models like GPT-4o.

The results indicated that GPT-4o exhibits a moderate level of lexical diversity and syntactic complexity, with certain chapters demonstrating higher readability and sentence complexity, while others favored simpler structures. The consistency in word frequency and stopword count across chapters suggests a relatively formulaic approach to text generation, a hallmark of machine-generated content. Additionally, the NER analysis revealed notable thematic variations between chapters, particularly in the prevalence of entities such as PERSON, ORG, and GPE, offering insights into the model's ability to handle different contextual content.

While the analyses did not detect significant variations in the syntactic or lexical features across chapters, the combination of metrics suggests that GPT-4o maintains a consistent style while adapting to different content areas. This consistency is a critical characteristic of machine-generated texts and highlights the potential for identifying machine-authored content through stylometry.

### 5.1. Future Directions

Several areas of research can be pursued based on the findings of this study:

- **Deepening Stylometric Analysis:** While this study provided insights into surface level linguistic features, future work could incorporate more advanced stylometric techniques, such as n-gram analysis, character-level features, or deep learning models designed for authorship attribution. These methods could provide a more nuanced understanding of machine-generated text and its distinguishing characteristics from human writing.

- **Cross-Model Comparison:** A valuable extension of this research would involve comparing the stylistic fingerprints of different generative models (e.g., GPT-3, GPT4o, ChatGPT) to explore variations in linguistic and stylistic outputs. This could aid in better understanding how models evolve over time and how their text generation styles differ.

- **Domain-Specific Stylometry:** Exploring the linguistic fingerprints of machine generated texts in specific domains, such as legal, medical, or literary writing, could help uncover more domain-specific features. This could be particularly valuable for identifying machine-generated text in niche fields where certain stylistic norms are strictly adhered to.

- **Improving Authorship Detection Algorithms:** Future research could focus on developing and refining authorship detection algorithms that specifically target machine generated content. These algorithms could help differentiate between human and machine authorship with higher accuracy, providing useful tools in the fight against plagiarism and ensuring content authenticity.

In conclusion, this study has provided a foundational analysis of GPT-4o-generated texts, revealing consistent stylistic tendencies that reflect the capabilities and limitations of the model. By continuing to develop and refine stylometric analysis techniques, future research can further improve

our ability to identify machine-generated texts, understand their underlying structure, and address the ethical challenges associated with their increasing use.

**Acknowledgment**

I would like to express my deepest gratitude to my daughter, Yasmine Abu Adla, for her invaluable assistance with the statistical analysis in this study. Her support and contribution have been instrumental in the successful completion of this research.

**References**

Abbasi, A., & Chen, H. (2006, May). Visualizing authorship for identification. In *International Conference on Intelligence and Security Informatics* 60-71. Springer Berlin Heidelberg.

Abbasi, A., & Chen, H. (2008). Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS)*, 26(2), 1-29.

Banat, M., & Adla, Y. A. (2023). Exploring the effectiveness of GPT-3 in translating specialized religious text from Arabic to English: A comparative study with human translation. *Journal of Translation and Language Studies*, *4*(2), 1-23.

Delcourt, C. (2002). Stylometry. *Revue belge de philologie et d'histoire*, 80(3), 979-1002.

Diwan, N., Chakravorty, T., & Shafiq, Z. (2021). Fingerprinting fine-tuned language models in the wild. *arXiv preprint* arXiv:2106.01703.

Farshoukh, M. (2018). *Soul breezes*. Beirut: Iijaz Forum.

Iqbal, M. M., Raza, A., Aslam, M. M., Farhan, M., & Yaseen, S. (2023). A stylometric fingerprinting method for author identification using machine learning. *Technical Journal*, *28*(01), 28-35.

Islam, R., & Moushi, O. M. (2024). GPT-4o: The cutting-edge advancement in multimodal LLM. *Authorea Preprints*.

Jiao, W., Wang, W., Huang, J. T., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv preprint* arXiv:2301.08745.

Kocmi, T., & Federmann, C. (2023). Large language models are state-of-the-art evaluators of translation quality. *arXiv preprint* arXiv:2302.14520.

Kumarage, T., & Liu, H. (2023, November). Neural authorship attribution: Stylometric analysis on large language models. In *2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)* 51-54. IEEE.

Neal, T., Sundararajan, K., Fatima, A., Yan, Y., Xiang, Y., & Woodard, D. (2017). Surveying stylometry techniques and applications. *ACM Computing Surveys (CSuR)*, 50(6), 1-36.

Niess, G., & Kern, R. (2024). Stylometric watermarks for large language models. *arXiv preprint* arXiv:2405.08400.

Pavelec, D., Oliveira, L. S., Justino, E. J., & Batista, L. V. (2008). Using conjunctions and adverbs for author verification. *J. Univers. Comput. Sci.*, 14(18), 2967-2981.

Ramnial, H., Panchoo, S., & Pudaruth, S. (2016). Authorship attribution using stylometry and machine learning techniques. In *Intelligent Systems Technologies and Applications: 1* (113-125). Springer International Publishing.

Raunak, V., Sharaf, A., Wang, Y., Awadallah, H. H., & Menezes, A. (2023). Leveraging GPT-4 for automatic translation post-editing. *arXiv preprint* arXiv:2305.14878.

Son, J., & Kim, B. (2023). Translation performance from the user's perspective of large language models and neural machine translation systems. *Information*, 14(10), 574.

Stańczyk, U. (2010). DRSA decision algorithm analysis in stylometric processing of literary texts. In *Rough Sets and Current Trends in Computing: 7th International Conference, RSCTC 2010, Warsaw, Poland, June 28-30, 2010. Proceedings 7* 600-609. Springer Berlin Heidelberg.

Xu, J., Wang, F., Ma, M. D., Koh, P. W., Xiao, C., & Chen, M. (2024). Instructional fingerprinting of large language models. *arXiv preprint* arXiv:2401.12255.

Yan, J., Yan, P., Chen, Y., Li, J., Zhu, X., & Zhang, Y. (2024). GPT-4 vs. human translators: A comprehensive evaluation of translation quality across languages, domains, and expertise levels. *arXiv preprint* arXiv:2407.03658.

Zaitsu, W., & Jin, M. (2023). Distinguishing ChatGPT (-3.5,-4)-generated and human-written papers through Japanese stylometric analysis. *PLoS One*, *18*(8), e0288453.
Zaitsu, W., & Jin, M. (2023). Distinguishing ChatGPT (-3.5,-4)-generated and human-written papers through Japanese stylometric analysis. *PloS One*, *18*(8), e0288453.
Zhu, W., Liu, H., Dong, Q., Xu, J., Huang, S., Kong, L., ... & Li, L. (2023). Multilingual machine translation with large language models: Empirical results and analysis. *arXiv preprint* arXiv:2304.04675.