Harmony in Text

Exploring the Interplay between Musical Preferences and Language

Eliran Shem Tov

A thesis submitted for partial fulfillment for
Master's in computer science

Supervisor: Dr. Ella Rabinovich

School of Computer Science
The Academic College of Tel Aviv-Yaffo
Tel-Aviv, Israel

September 2024

Abstract

This thesis examines the complex interplay between personality traits and musical preferences, utilizing advanced natural language processing (NLP) techniques and cutting-edge large language models (LLMs) to generate and analyze textual content. Grounded in the Big Five personality traits model—openness, conscientiousness, extraversion, agreeableness, and neuroticism—the study explores the relationship between a person's musical preferences and their personality as it is manifested through their spontaneous, non-music-related, linguistic production on social media. A large-scale and diverse dataset comprising 575,816 text samples from 5,000 individuals with distinct musical preferences was curated from non-music-related online forums to ensure unbiased extraction of personality traits.

The research employed LLMs to generate a unique dataset labeled with personality traits, subsequently training logistic regression models to predict these traits from textual content. These models were applied to extract personality from the curated large-scale dataset and statistical analyses of the results, including ANOVA and t-tests, were conducted to identify significant correlations between the extracted personality traits and musical preferences. The results revealed that Classical music enthusiasts exhibited higher agreeableness and lower extroversion, Hip-Hop fans showed lower agreeableness and higher neuroticism, Metal fans showed a higher propensity for neuroticism, and electronic music listeners demonstrated greater openness. These findings highlight the complex interplay between personality traits and musical tastes, suggesting that certain personality characteristics are strongly linked to specific music genres. The study contributes to the growing body of research at the intersection of psychology and computational linguistics, shedding some new and interesting light on the role of personality in shaping, and being shaped by, musical preferences.

Content

1. Introduction

"**Our language is the reflection of ourselves. A language is an exact reflection of the character and growth of its speakers.**" – Cesar Chavez [1]. This sentiment underscores the profound connection between self-expression and identity, a theme that is deeply mirrored in music. Music is a powerful expression of individuality, often reflecting deeper aspects of one's personality. The connection between musical preferences and personality traits has long been a subject of interest in psychology. The Big Five personality traits model—comprising openness, conscientiousness, extraversion, agreeableness, and neuroticism—provides a robust framework for exploring this relationship. Prior research indicates that certain genres are associated with specific personality traits, suggesting that music genre preferences can offer valuable insights into an individual's psychological profile.

Recent advancements in natural language processing (NLP) and large language models (LLMs) have opened new avenues for personality assessment, particularly through the analysis of text. By examining language patterns, textual expression, and stylistic choices, advanced NLP methods have the potential to predict the presence of personality traits, though existing methods often struggle with low accuracy when tested on unseen data. One of the contributions of this work is the development of a (more) powerful and accurate classifier, made possible by leveraging high-quality and diverse training data produced by contemporary LLMs. Along with this improvement, evaluating the linguistic production of subjects offers a scalable, unbiased, and non-intrusive approach to understanding personality, utilizing the richness of human language as a proxy for deeper psychological characteristics.

This thesis seeks to investigate the relationship between personality traits, as extracted from text, and musical preferences. We study whether and how a person's musical preferences are manifested in their authentic written linguistic production. Specifically, we ask **if there are detectable (similarities and) differences in the Big Five personality traits in the language of people with various musical preferences when authoring spontaneous textual content on social media**. The hypothesis driving this research posits that despite the inherent complexity of human personality and musical tastes, the differences in personality traits of people with various musical preferences, "shine through" their language to the extent that can be automatically captured through advanced NLP techniques.

To explore this hypothesis, a large and diverse dataset was curated, consisting of text samples from individuals with reliably identified musical preferences. These samples were collected from a range of public non-music-related online forums (Sub-Reddits), ensuring a broad topical representation. Each participant was linked to a single dominant music genre they frequently engage with, enabling unbiased textual personality extraction and a focused analysis of the relationship between personality traits and musical preferences.

The methodology of this research encompassed several key steps: (a) curating a large-scale dataset consisting of 575,816 texts authored by 5,000 individuals with distinct musical preferences; (b) evaluating existing tools for personality trait extraction; (c) utilizing large language models (LLMs) to generate a unique dataset of texts labeled with personality traits and training logistic regression classifiers to predict these traits based on textual content; (d) extracting personality traits from the curated dataset mentioned in step (a); and (e) conducting statistical analyses to identify significant differences in the extracted personality traits between fans of different music genres. This multi-faceted approach enabled a thorough exploration of the research question.

The results of this study revealed significant associations between musical preferences and personality traits, as extracted from textual content. **Classical** music enthusiasts were found to exhibit **higher levels of agreeableness** and **lower levels of extroversion**, while **Hip-Hop** fans displayed **lower agreeableness** and **higher neuroticism**. Additionally, **Metal** fans also showed a **higher propensity for neuroticism**, and **electronic** music listeners demonstrated **greater openness**. These findings corroborate, to some extent, evidence previously shown for musical preferences relations with self-reported personality traits but goes above and beyond by demonstrating that musical preferences are manifested in language through personality traits. Supported by multiple statistical methods, including ANOVA and T-Tests, the findings confirmed the statistical and practical significance of these relationships, offering valuable insights into how personality may shape, and be shaped by musical preferences.

2. Background

Music has long captivated humanity, serving as a powerful medium for emotional expression, cultural exchange, and individual identity formation. Research has extensively explored the psychological and neurological underpinnings of our connection to music, demonstrating its ability to influence mood, memory, and cognitive processes. A growing body of work delves deeper, investigating the interplay between musical preferences and various aspects of an individual, including their personality.

This study contributes to this ongoing exploration by examining the relationship between musical preferences and personality through linguistic production. Prior research has established that language offers valuable insights into personality, with specific word choices, writing styles, and emotional tones reflecting individual differences. Building on this foundation, we explore whether and how a person's personality traits are reflected in their spontaneous written language and if the relation to musical preferences is significant.

The notion that language can reveal personality traits has roots dating back to early psychology. Modern research has adopted a data-driven approach, leveraging advancements in natural language processing (NLP) to analyze textual features and their correlations with personality dimensions. The Big Five personality framework, encompassing Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN), serves as a widely used model for personality assessment, along with other models. Studies have identified links between specific linguistic features and these traits.

The link between musical preferences and personality has been a subject of ongoing investigation in psychology. Early research explored the potential of using music preferences as indicators of mental health. With the emergence of the Big Five model, studies have examined correlations between personality traits and preferred musical genres. While some studies have identified tentative associations (e.g., Openness with complex musical styles, Extraversion with high-energy music), the findings remain inconclusive. This complexity likely stems from the multifaceted nature of both music and personality. Musical preferences are influenced by cultural background, social influences, and personal experiences, while personality traits interact with each other in nuanced ways.

The explosion of social media platforms has yielded a wealth of user-generated textual data, providing researchers with a valuable resource for studying language and personality. Social media platforms allow us to observe individuals' natural writing styles in diverse contexts, offering a more authentic representation of language use compared to traditional self-reported measures. Furthermore, advancements in NLP techniques like word embeddings and deep learning models enable the automated extraction of meaningful features from large textual datasets. These developments empower researchers to analyze vast amounts of social media data and textual content in the wild and uncover subtle linguistic patterns associated with personality traits.

Despite the established links between language and personality, and the growing understanding of the music-personality relationship, the intersection of these domains remains relatively unexplored. Limited research has investigated whether musical preferences leave detectable traces in an individual's written language style. This study aims to bridge this gap by leveraging the power of social media and NLP to explore how musical preferences reflect personality types in spontaneous written language, and how significant those relationships are. By analyzing the language used by individuals with distinct musical preferences, we seek to identify characteristic patterns and shed light on the potential connections between music, personality, and the way we express ourselves in writing.

2.1. Human Personality Analysis

Human personality analysis is a crucial field in psychology and behavioral sciences that aims to understand and categorize individual differences in thoughts, emotions, and behaviors. This analysis is essential for various applications, including mental health treatment, educational strategies, and organizational management. Personality traits are generally considered stable over time and significantly influence how individuals interact with their environment and other people. Several models have been developed to explain and measure these traits, with the Big Five personality traits being one of the most widely accepted frameworks.

2.1.1. The Big Five Traits Model

The Big Five personality traits provide a comprehensive taxonomy for understanding human personality. These traits include Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, often remembered by the acronym OCEAN. Each trait encompasses a spectrum, where individuals may score high or low, influencing their behavior and interaction patterns.

1. **Openness (to Experience)**: This trait features characteristics such as imagination, creativity, and a broad range of interests. Individuals high in openness are often more adventurous and open to new experiences and ideas. Studies have shown that openness is linked to higher creativity and a preference for novelty [2].
2. **Conscientiousness**: This trait is associated with high levels of thoughtfulness, good impulse control, and goal-directed behaviors. Highly conscientious individuals are organized and mindful of details. Conscientiousness has been shown to be a strong predictor of academic and job performance [3][4].

3. **Extraversion**: This trait includes characteristics such as excitability, sociability, talkativeness, assertiveness, and high amounts of emotional expressiveness. Extroverted individuals often enjoy being with people, participating in social gatherings, and are perceived as full of energy. Extraversion has been linked to positive social interactions and higher life satisfaction [5].
4. **Agreeableness**: This trait reflects individual differences in general concern for social harmony. Agreeable individuals value getting along with others and are generally considerate, kind, generous, and willing to compromise their interests with others. High agreeableness is associated with better social interactions and reduced conflict [6].
5. **Neuroticism**: This trait is characterized by sadness, moodiness, and emotional instability. Individuals high in neuroticism tend to experience mood swings, anxiety, irritability, and sadness. Neuroticism is linked to poorer mental health outcomes and increased risk of various psychological disorders [5].

The Big Five model provides a robust framework for understanding personality, which has been validated across cultures and contexts [2]. The traits have significant implications for predicting a wide range of behaviors and outcomes in personal, social, and professional domains.

2.1.2. Other Personality Models

The Big Five Traits Model is predominant; however, several other models offer alternative perspectives on personality, emphasizing other aspects and dimensions:

- **HEXACO Model**: This model adds a sixth dimension, Honesty-Humility, to the Big Five traits, capturing traits like fairness, sincerity, and modesty, which are not fully accounted for in the traditional Big Five.
- **Alternative Five Model**: Proposed by Zuckerman, this model includes five factors: Impulsivity, Sociability, Aggression, Activity, and Sensation-seeking, focusing more on temperamental and behavioral traits.
- **The Dark Tetrad Model**: This model focuses on four negative personality traits: Machiavellianism, Narcissism, Psychopathy, and Sadism. It is particularly useful in understanding antisocial behaviors and darker aspects of personality.
- **Supernumerary Traits Model**: This model suggests the existence of additional personality traits beyond the Big Five, emphasizing unique individual differences that may not be captured by the conventional dimensions.

While the Big Five Traits Model remains a cornerstone in personality psychology, these alternative models provide valuable insights by highlighting different dimensions and aspects of personality. The HEXACO model expands the framework to include moral character, the Alternative Five Model emphasizes temperamental and behavioral traits, the Dark Tetrad focuses on darker, antisocial aspects, and the Supernumerary Traits Model underscores the uniqueness of individual differences. Together, these models enrich our understanding of the complex and multifaceted nature of human personality.

2.2. Personality and Musical Preferences

Music, a universal aspect of human culture, has the power to reflect and influence an individual's personality. Research has increasingly explored the link between personality traits and musical preferences, suggesting that the music one prefers can offer insights into their personality. Understanding these relationships can have practical applications in areas such as marketing, therapy, and social interactions.

2.2.1. Influence of Personality Traits on Music Preferences

Personality traits significantly shape our preferences for different musical styles. The Big Five personality traits have been particularly useful in predicting these preferences.

1. **Openness**: Individuals high in openness tend to prefer more complex and novel music genres, such as jazz and classical music. This preference is linked to their appreciation for creativity and variety [7].
2. **Conscientiousness**: Those high in conscientiousness often prefer structured and conventional music styles. These individuals are likely to enjoy genres that reflect order and discipline, such as pop and classical music [8].
3. **Extraversion**: Extraverts are inclined towards energetic and rhythmic music. They often enjoy upbeat and lively genres such as electronic dance music (EDM), pop, and hip-hop, which align with their sociable and active nature [9].
4. **Agreeableness**: Individuals who score high on agreeableness tend to prefer mellow and harmonious music. They often enjoy genres like folk, country, and soft rock, which are perceived as warm and soothing [10].
5. **Neuroticism**: Those high in neuroticism may prefer intense and emotional music genres. They often listen to genres like heavy metal, punk, and emo, which resonate with their emotional intensity and variability [11].

2.2.2. Empirical Evidence

A growing body of empirical research supports the connection between personality traits and musical preferences:

- **Meta-analyses and large-scale studies**: Meta-analyses have revealed consistent yet modest correlations between personality traits and music preferences. For instance, openness has shown small but significant correlations with preferences for sophisticated musical styles [7]. Similarly, extraversion has been linked to preferences for energetic and rhythmic music [9].
- **Behavioral studies**: Research using behavioral data, such as active listening habits and social media interactions, has further validated these findings. For example, studies using Facebook Likes and online music streaming data have demonstrated that preferences for certain genres can reliably predict personality traits [9].

- **Cross-cultural studies**: Research across different cultures has shown that while specific musical preferences may vary, the underlying associations between personality traits and music preferences remain consistent. This suggests a universal aspect to the way personality influences musical tastes [12].

In conclusion, the interplay between personality traits and musical preferences offers a fascinating glimpse into how deeply intertwined our personal characteristics are with our artistic choices. By leveraging advancements in data analysis and natural language processing, researchers can continue to uncover the subtle nuances of these relationships, providing deeper insights into human behavior and preferences.

3. Related Work

3.1. Personality Detection

The extraction of personality traits from textual data has garnered significant interest over the years due to its potential applications in psychology, human-computer interaction, and personalized services. The trends and advancements in deep learning-based personality detection emphasize its diverse applications, including enhanced personal assistants, recommendation systems, word polarity detection, specialized healthcare and counseling, forensics, job screening, psychological studies, and political forecasting. A comprehensive review [13] by Mehta et al. [13] underscores the effectiveness of integrating multiple modalities, such as text, audio, and visual content, to improve personality extraction processes, with late fusion being a common technique for combining these sources.

Acceptance issues related to personality-based recommender systems are addressed through empirical user studies, demonstrating that incorporating personality information can significantly enhance user loyalty and reduce cognitive effort. These systems are perceived as more useful and easier to use, leading to higher overall user satisfaction [14]. Personality-based personalization could come in handy also in the dining, tourism, and hospitality industries. Large-scale research has employed BERT (a Bidirectional Encoder Representation from Transformers, i.e. pretrained large language model that can capture semantic meaning within text) to classify personality based on texts from a tourism-related dataset. It showed that integrating personality traits into the hospitality and dining recommendation process significantly improved the accuracy and relevance of user suggestions [15].

In addition, personalizing music recommendations based on personality traits can significantly enhance user experience, particularly when tailored to specific age groups [16]. However, the implications of musical preferences and personality relationships aren't limited to traditional recommendation systems. The potential of big data in understanding human behavior related to music and its impact on health suggests that streaming services could leverage personality data to provide personalized health and medical solutions [17].

Tupes and Christal's (1961) [18] research was among the first to identify the five broad personality factors. Their analysis of personality data from Air Force personnel revealed consistent patterns that later became the basis of the Big Five traits, which have since become a cornerstone for personality assessment methods [18].

Following the foundation laid by earlier works, researchers like Goldberg further refined the Big Five model in 1981, emphasizing its robustness and replicability, which helped establish it as a dominant model in personality psychology [19]. Later work provided a foundational understanding of the Big Five personality traits, and offered a thorough review of measurement techniques, and theoretical implications. This work has significantly contributed to the widespread acceptance and use of the Big Five taxonomy in both research and applied psychology [20].

3.2. Identifying Personality from Text

The concept of measuring personality through linguistic analysis can be traced back to the pioneering work of Francis Galton in 1884. Galton hypothesized that personality could be effectively captured through the adjectives found in language and written text, laying the groundwork for future studies in personality extraction [21]. Pennebaker and King explored linguistic styles as individual differences, confirming correlations between language use and personality traits [22].

The introduction of the Linguistic Inquiry and Word Count (LIWC) method by Pennebaker et al. in 1996 marked a substantial step forward. LIWC provides a statistical analysis of text to determine the emotional and cognitive status of the writer, establishing itself as a crucial tool for text-based personality assessment [23].

Advancements in computational methods have significantly improved the accuracy of text-based personality detection. Mairesse et al. (2007) developed methods using linguistic cues to automatically recognize personality traits in both conversation and text, focusing on the Big Five personality traits [24]. Around that time, social networks such as Twitter, Facebook, and Reddit emerged and significantly affected our lives, including the way we express ourselves textually. The great accessibility and presence of virtual keyboards and social platforms, wherever we go, introduced us to a new era of highly available big data in the wild, textual and non-textual. In 2007, Kosinsky et al. generated the MyPersonality dataset which was later [25] published along with a research paper that highly focused on exposing the privacy risks related to the users' data collected online, emphasizing the sensitivity of textual content they share and what it could tell about them. MyPersonality was a Facebook application that allowed users to self-report and fill in a personality questionnaire. Six million people used it and almost 2.5 million users chose to contribute their Facebook data, including their posts. The MyPersonality dataset is the set of users' unsanitized posts with their self-reported personality labels. Since 2018, only a partial dataset of 10K posts, made by 250 users is available for research. This paper's findings informed new EU and U.S. privacy laws and Facebook privacy policies. Since then, dozens of research works have utilized social media datasets, and mostly this one and subsets of it [26], to explore the ability to extract personality traits from textual content. Based on the study [25], Cambridge University's Psychometrics Centre has published an online application for research purposes, called Apply Magic Sauce. It allows users to upload textual content from Twitter, Facebook, LinkedIn, and open text and get a detailed prediction of their personality by the Big Five model.

Apply Magic Sauce is one of the publicly available tools we used in our work, to generate a personality prediction benchmark.

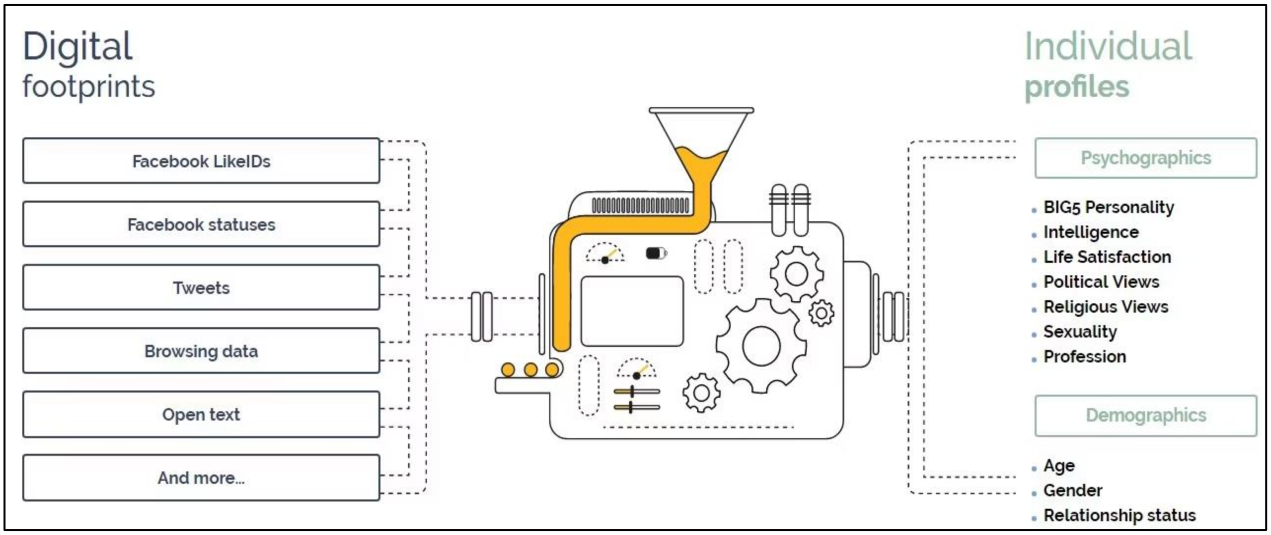

Figure 1: Apply magic sauce illustration

Other works like Quercia et. al [27] have mined Twitter usage patterns of users and their tweets to predict their personalities. Wang [28] used a large-scale dataset of 90K Twitter users and applied NLP techniques like "bag of n-grams", POS tags, and word vectors, and showed an average accuracy of 0.66 in predicting personality traits. Bawa et al. who also utilized the MyPersonality dataset, demonstrated the extraction of personality traits from written comments using machine learning techniques, further advancing the field[29].

Recent advancements in deep learning have further enhanced the capabilities of personality detection from text. Majumder et al. utilized deep learning-based document modeling to accurately detect the Big Five personality traits, outperforming state-of-the-art methods for all five traits [30]. Sun et al. proposed a model combining bidirectional LSTMs with CNNs for personality detection, demonstrating significant improvements in accuracy [31]. Others trained semi-supervised learning algorithms, concluded by LIWC, N-Gram along with Word2Vec word embedding, on a social media dataset to predict personality through text [32]. Ren et al. integrated semantic and emotional features using BERT to vectorize texts at a sentence level, utilizing the self-attention mechanism. Concatenation of those vectors with the sentiment vectors showcased the potential of combining different feature sets for improved performance [33].

The field has continued to evolve with recent contributions that integrate new methodologies and expand upon previous research. For instance, Kosan et al. [26] developed a method for predicting personality traits using semantic structures and LSTM-based neural networks. This study utilized the IBM Personality Insight service and the PAN-2015 Personality Dataset to enhance the accuracy of personality trait predictions [26].

Various deep learning approaches were proposed [34] for text-based personality prediction using multiple data sources, including the Essays and myPersonality datasets. Their method demonstrated significant improvements in the prediction accuracy of the Big Five traits. Yang et al. have employed the Pandora dataset, which is a large-scale collection of 3,000,566 Reddit comments from 1,568 users and their corresponding personality traits elicited using surveys involving the Big-Five constructs, along with known datasets like MyPersonality and others. It introduced a deep learning artifact for text-based measurement of personality, leveraging concepts from both psycholinguistic theories and advanced deep learning strategies. Their model, which is not publicly available online, incorporated transfer learning and hierarchical attention networks to improve personality detection accuracy [35].

The advent of large language models (LLMs) has revolutionized the field of personality extraction from text, offering unprecedented accuracy and scalability. Recent research by Peters and Matz tested whether GPT-3.5 and GPT-4 could derive the Big Five personality traits from users' Facebook status updates, taken from the MyPersonality dataset, in a zero-shot learning scenario. Their findings showed an average correlation of r = .29 between LLM-inferred and self-reported trait scores, a level of accuracy that is slightly lower than that accomplished by supervised models, trained or fine-tuned for this purpose and with the same data (e.g., Park et al. [36], who reported correlation average r=0.37). This study underscores the potential of LLMs in personality extraction, particularly in the context of social media data [37]. Another significant advancement comes from the work of Safdari et al. has explored the measurement and shaping of personality traits within the outputs of large language models. This study presented a comprehensive method for administering and validating personality tests on widely used LLMs and demonstrated that personality measurements in the outputs of some LLMs are reliable and valid under specific prompting configurations. Furthermore, it was found that the reliability and validity of synthetic LLM personalities are stronger in larger and instruction fine-tuned models. The study also discussed the potential for shaping personality traits in LLM outputs to mimic specific personality profiles, highlighting the application and ethical implications of these methods [38].

3.3. Personality and Music Preferences

The relationship between musical preferences and personality traits has been a significant focus of academic inquiry for many years as well. The exploration of this link began with the work of Raymond Cattell in the 1950s. Cattell and his colleagues suggested that music could satisfy deep and unconscious needs, thus providing insights into personality. They developed the IPAT Music Preference Test, which identified stable music-preference factors reflecting unconscious aspects of personality [39][40]. Research on the Big Five personality traits has significantly influenced studies on musical preferences. Rentfrow and Gosling found that Openness to Experience correlated with a preference for complex and novel music, such as classical and jazz, while Extraversion was linked to energetic and rhythmic music. Agreeableness was associated with upbeat and conventional music [41]. Bonneville-Roussy et al. extended these findings by analyzing trends in musical engagement and preferences from adolescence through middle adulthood, showing that Openness to Experience remains a strong predictor for preference for complex music across different age groups [42]. A study by Ferwerda et al. analyzed relationships between music listening history from Last.FM users and the myPersonality dataset. They conformed with earlier studies and found that the Openness to Experience trait is correlated with many music genres, including new age, classical, world, blues, country, folk, jazz, and alternative. Samples with a high score in Conscientiousness showed a negative correlation with folk and alternative music, while Extraversion correlated with R&B and rap music. Agreeableness positively correlated with country and folk, and Neuroticism only showed a positive correlation with alternative music [16].

Rentfrow et al. aimed at refining the understanding of these relationships and introduced the MUSIC model, a genre-free taxonomy that reflects emotional and affective responses to music, linking each factor to the Big Five personality traits [43]. This model has been pivotal in providing a common language for researchers.

A study by Greenberg et al. (2022) examined musical preferences across 53 countries, revealing that Extraversion was correlated with stronger reactions to contemporary musical styles, whereas Openness was correlated with sophisticated musical styles. Neuroticism showed a preference for intense musical styles, reflecting inner angst and frustration [44].

Advancements in computational methods have enabled large-scale studies of the relationship between music and personality. For instance, Nave et al. demonstrated that musical preferences and behaviors, such as Facebook likes for musical artists, can predict personality traits, showing the effectiveness of combining social media data with traditional psychological assessments [9]. Anderson et al. linked Spotify music listening data to personality traits, demonstrating that certain musical behaviors could predict traits such as Openness and Extraversion. This study highlighted the potential of streaming data in personality research [45].

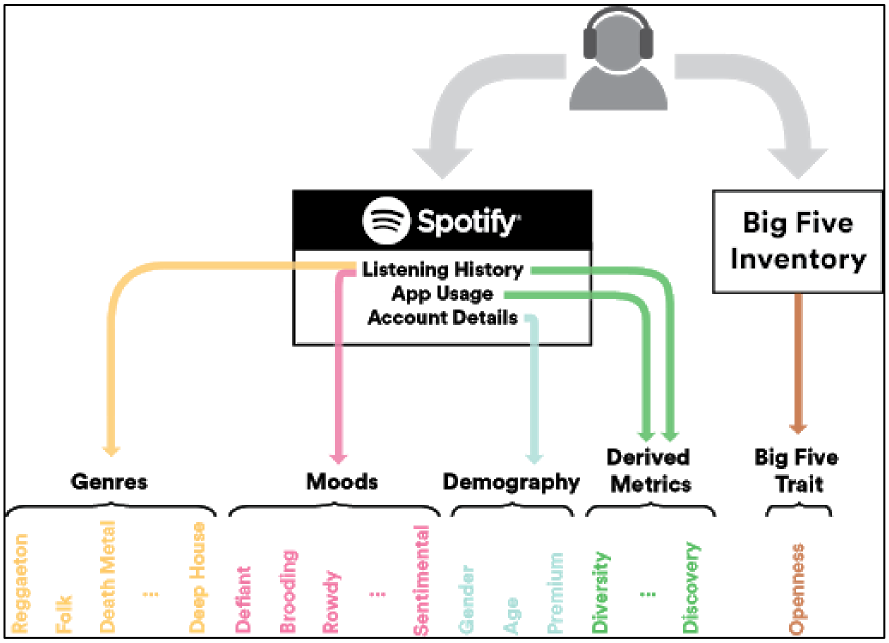

Figure 2: Illustration of Spotify's research that established external validity about the musical preferences and Big Five

The field has continued to evolve with recent contributions integrating new methodologies. A study by Sust et al. proposed a personality computing approach, quantifying participants' music preferences via audio and lyric characteristics of their played songs. This study used technical audio features from Spotify and textual attributes obtained via natural language processing to predict Big Five personality traits on domain and facet levels [46].

3.4. Text, Personality and Musical Preferences

The exploration of personality extraction and its intersection with musical preferences reveals a multifaceted and evolving field. Early foundations laid by researchers like the abovementioned Galton [21] and the development of the Big Five model have provided a robust framework for understanding personality traits. Advancements in computational methods, particularly the use of deep learning and large language models, have significantly enhanced the accuracy and scalability of personality detection from text. Simultaneously, studies on musical preferences have demonstrated consistent correlations between the Big Five personality traits and specific music genres, for instance, individuals high in Openness tend to prefer classical and jazz music, and often enjoy a wide variety of unconventional music styles. Neuroticism on the other hand correlates positively with a preference for conventional popular music, which may indicate that those scoring high on Neuroticism use such music to mitigate negative effects induced by more arousing music [47].

The combination of these domains offers a comprehensive understanding of how individual traits manifest across different contexts. In some ways, the two fields complete each other and allow personality prediction based on both musical preferences and textual content, as well as highlighting the interplay between individual preferences in general and personality traits. With contemporary research leveraging social media and streaming data for large-scale analysis, this integrated approach not only enriches personality research but also opens new avenues for personalized applications in various fields, from psychological assessments and personalized health care to targeted marketing and user experience design.

3.5. Conclusion

Most *personality extraction from text* studies relies on a limited number of datasets, such as the Essays and myPersonality datasets, which can constrain the diversity and generalizability of findings. Additionally, these studies predominantly use self-reports and surveys to gather personality data, which limits their credibility and scalability. The requirement for participants to fill out detailed surveys restricts the potential for large-scale data collection and introduces biases related to self-perception and response accuracy. Addressing these issues is crucial for advancing the field and achieving more robust, generalizable insights.

While many prior findings are suggestive that there is a strong relationship between musical preferences and personality traits, and others have established the idea that personality traits can be extracted from text with high accuracy, only a limited body of research has examined the intersection between the two. For instance, Neuman et. al [48] explored the hypotheses that personalities are associated with lyrics of various music genres and that personality as expressed in song lyrics may be used for genre classification. However, no studies have directly addressed the connection between these two as our question aims to address it: Does our personality, as it “shines through” our writing, predict our musical preferences?

4. High-level Overview of the Methodology and Experimental Approach

As we mentioned, textual content can provide cues about the writer's personality and enable personality trait extraction by several models. Similarly, the existing literature shows a solid correlation between personality traits and musical preferences.
The related work in these two fields has constantly shown a few repeating caveats:

1. Self-reports for musical preferences:
    a. May reflect built-in biases
    b. May hide a lack of musical knowledge or understanding, leading to misclassification of musical preferences
    c. Doesn't scale, as they require filling manual or online questioners
2. Personality classification questionnaires:
    a. May reflect built-in biases
    b. Individual sample classification cannot be easily validated
    c. Doesn't scale, as they require filling manual or online questioners

To address these weaknesses, our experiment leveraged the abundance of textual content available online, particularly on social media platforms where users voluntarily publish their writing. This approach allows us to label the content with unique user identifiers, enabling more accurate data collection and analysis.

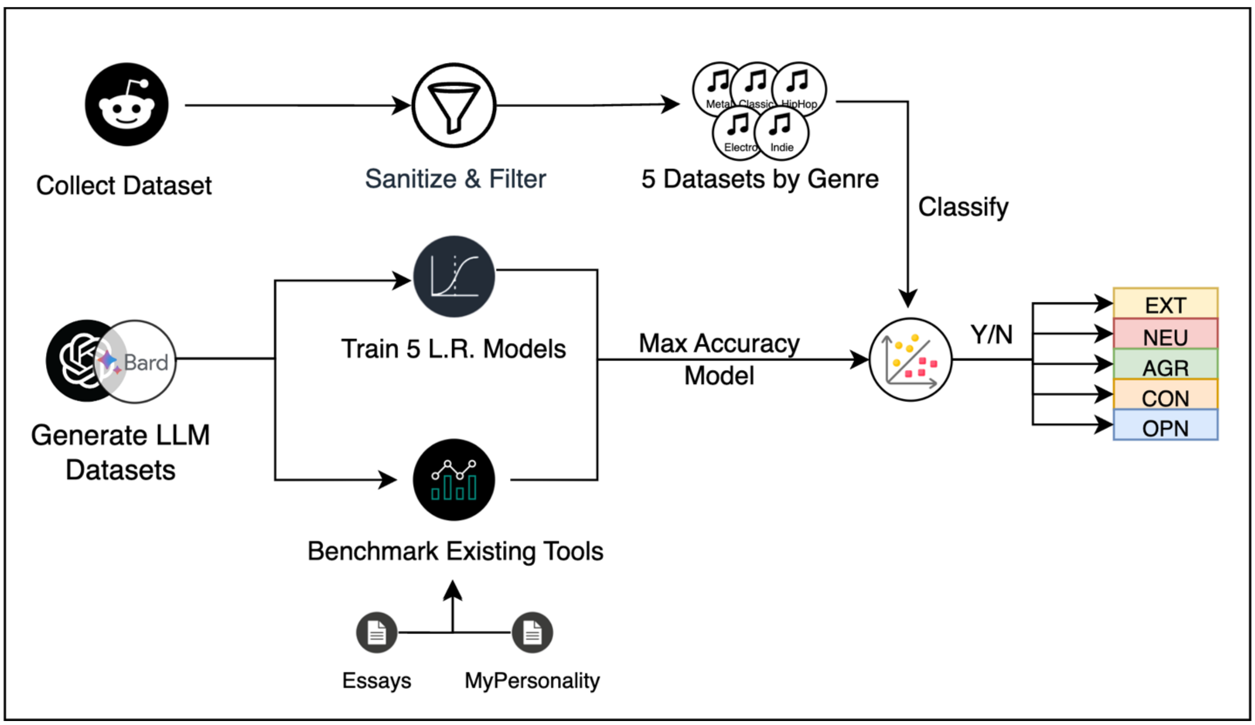

Figure 3: Experiment's Overall Approach

Reddit is a well-known social media platform [49] with over 130 thousand communities, also known as subreddits [50]. There are hundreds of music-related subreddits with millions of subscribers, some of which are genre-specific, making it a great source of topical textual data. As we describe in detail in the next section, we collected the activity in five distinct genre-related subreddits, namely Classical, Indie, Hip-Hop, Electronic, and Metal. We analyzed this data to extract the 1000 most active users in each that are not active in any other music-related subreddit and collected their activity from non-music related subreddits, resulting with 575,816 posts and comments that are not related to music, made by the 1000 most active users in each genre's subreddit.

Before we moved on with our freshly collected dataset, we tried to evaluate available tools for personality extraction from text by the Big Five traits model and form an accuracy baseline, detailed in section 7. To do so, we used the "gold standard" datasets for personality extraction: Essays and myPersonality, discussed in section 6. We generated predictions for each personality trait, i.e., extraversion, agreeableness, conscientiousness, neuroticism, and openness. As a common practice when evaluating classification models, the tools were tested for accuracy. Accuracy is calculated as the ratio of correctly predicted instances to the total instances. The procedure and accuracies achieved for the existing tools baseline on both datasets are fully detailed in the next sections.

The next step was to create our own, improved way to predict personality traits based on a given text. We treated the personality prediction problem as five distinct personality traits classification problems, and to address those we employed word-embedding and logistic regression techniques. Word embedding is a technique in natural language processing (NLP) that represents words as vectors in a continuous vector space. This allows words and phrases with similar meanings to have similar vector representations. Embeddings can capture the semantic relationships between words, making them excellent feature vectors for classification tasks such as ours. Using these embeddings, we trained a logistic regression model on the MyPersonality dataset and evaluated its performance on the Essays dataset. While the model's results were inconclusive, they outperformed the tools used to establish the accuracy baseline, as detailed in Table 3. Later analysis taught us that all the tools were at least partially trained on the datasets we used to form the accuracy baseline, a fact that had a potentially high impact on the predictions' accuracy of the baseline tools.

To overcome the limitations and biases of the "gold standard" training sets, as well as the drawbacks of existing tools, we leveraged the power of large language models (LLMs) and generative AI. Based on academic and online resources [58][59][60][61][62], we crafted detailed definitions for each of the Big Five personality traits [11]. Using these definitions, we generated consistent prompts 12] for the five personality traits. We utilized OpenAI's ChatGPT and Google's Bard (now Gemini) to create an unbiased and clear training set and test set for each personality trait consisted of hundreds of separate texts with both high and low presence of each trait.

Using the Chat-GPT generated training sets, we trained a logistic regression model for each personality trait and tested these models against the unseen test sets generated by Bard. As detailed in the upcoming sections, the models achieved high accuracy in predicting the presence or absence of personality traits in the unseen test dataset.

Once we had five relatively accurate logistic regression models, we came back to the data we collected from the five distinct musical genre subreddits, and after sanitizing it, we vectorized the 575,816 posts and comments, and employed the five logistic regression models to predict the probability of the presence of each trait for every user, based on its texts. As detailed in the results section, grouping the predictions by musical genre subreddit showed interesting and significant differences between the communities.

5. Data Collection

In this study, we sought to collect a large and diverse dataset of spontaneous and authentic written content from individuals with distinct musical preferences. We aimed to gather texts that reflect a wide range of commonly discussed topics, deliberately excluding any direct references to music. This approach was taken to avoid potential biases and to ensure that the analysis accurately captures the relationship between musical preferences and personality traits.

Social media platforms are a direct candidate to act as such data sources, and while there are a lot of famous social media platforms, each has its uniqueness. One of those is Reddit: a well-known, highly organized, and topic-categorized home of over 2.3 billion monthly active users [49], in over 130 thousand communities, also known as subreddits [50]. More than 500 subreddits have at least 1 million subscribers [51], including r/music with almost 33 million subscribers [52], as shown in Figure 4, which ranks the top twenty subreddits by number of subscribers (in millions). The dominant age group of Reddit users is 18 to 29, followed by 30 to 49 with 31%, while men account for almost 64% of the total user base (larger than the monthly active number of users) [53]. 72% of Reddit's users use the platform for entertainment purposes [54].

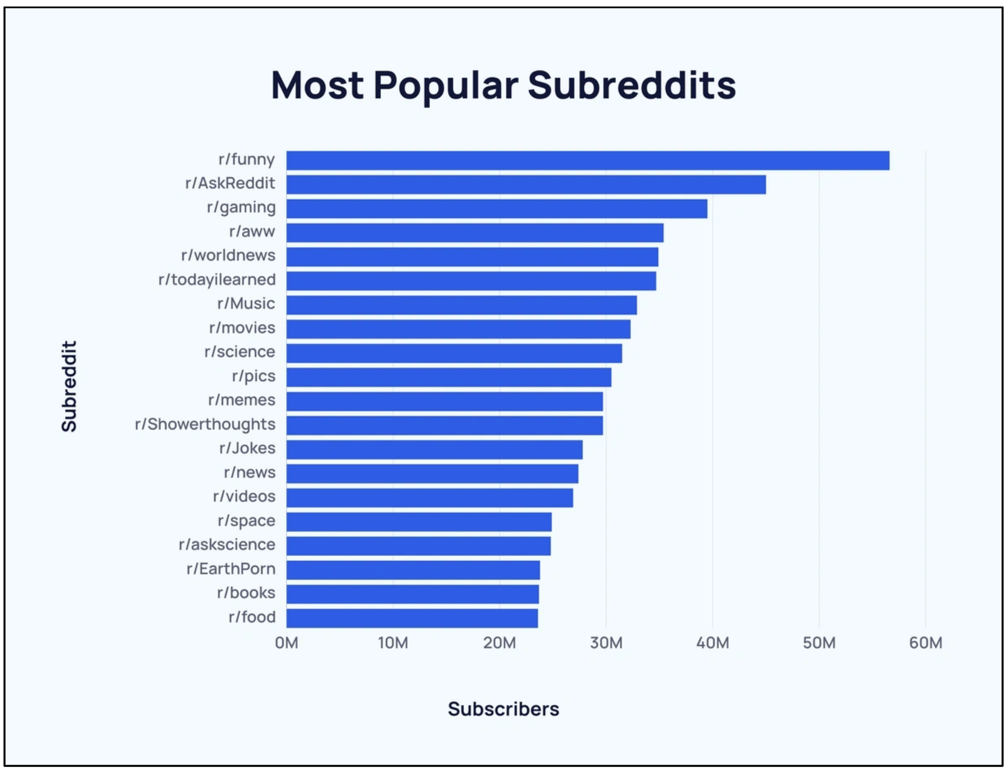

Figure 4: Each of the top 10 subreddits has more than 30 million subscribers [55].

5.1. Top Active Authors

After the initial validations, we decided to focus on five distinct large musical genre subreddits, namely Indie, Electronic, Hip-Hop, Metal, and Classical music. We extracted the most recently published ~1M posts and comments from each community and found that we can identify what we call “top active authors” per subreddit. For example, the fetched 1,000,154 comments in the Metal Music subreddit were all posted by 74,055 Redditors, let them be the Metal community’s top active authors.

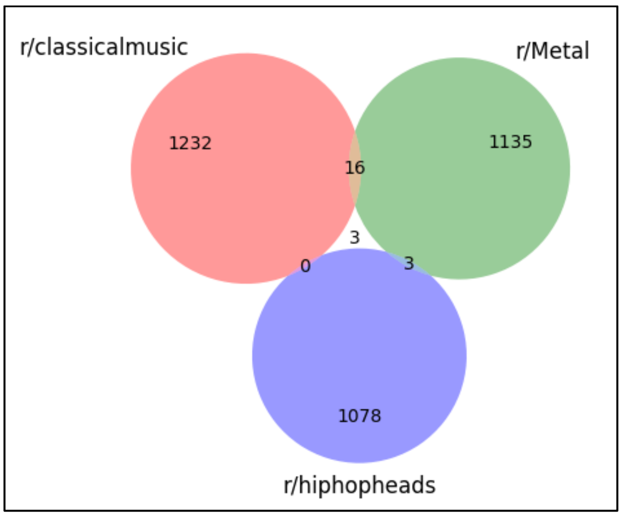

Figure 5: Example of intersection between different musical genres users

5.2. Authors with a single identifiable musical preference

Given the absence of validated preference annotations, we adopted an approximation method to identify users with distinct musical preferences. Specifically, we focused on users who (1) are subscribed to only one of the five selected musical genre communities, and (2) actively participate in discussions within those communities. Although this approach is not perfect, it is reasonable and aligns with similar methodologies employed in previous research [44]. The limitations of this approximation method will be discussed in the limitations section of the conclusions.

To ensure that the corpus we will work with is as unbiased and reliable as possible, we eliminated all the users that appear in any intersection with at least one of our five subreddits. Then we plotted the distribution of the number of comments for the author, per subreddit.

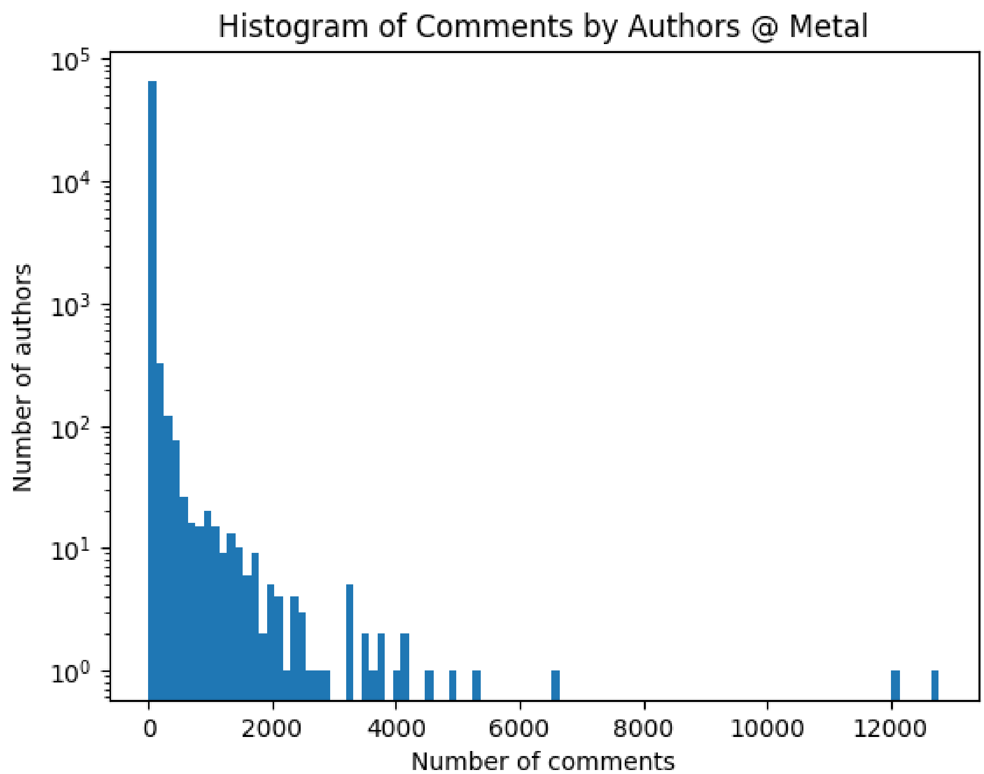

Figure 6: Number of comments for author distribution per subreddit, similar distributions were demonstrated in all 5 genres and could be explored as part of the code 9]

Community	500th	1000th	2000th
Indie	143	78	41
Electronic	111	70	44
Hip-Hop	145	87	51
Metal	179	88	45
Classical	191	109	58

A quick glance at the distribution of the number of author comments per subreddit, illustrated in figure 6, revealed a consistent pattern among our five selected subreddits. Based on this observation, we proceeded to rank the authors by the number of comments they had published and analyzed the number of comments contributed by the Nth author in each community, with N set at 500, 1,000, and 2,000.

Table 1: The number of comments for the 1000th commenter is close in all 5 subreddits, therefore we decided to focus on the top 1000 authors who are exclusively associated with each community.

5.2.1. Collecting diverse content written by authors from the identified communities

The next phase is collecting all the Reddit activity by the 5 communities' top 1000 authors with a single identifiable musical preference, from other subreddits, that is 5000 Redditors' activity excluding their activity in the 5 musical subreddits that we focused on. Given a user ID, the entire user's digital footprint can be retrieved from Reddit, allowing direct access to a person's linguistic productions across a variety of topics, and spanning multiple years.

5.2.2. Detailed Users' Activity – Initial Analysis

To better understand the nature of the collected data, we applied some exploration activities:

- For every musical genre (out of five): list the subreddits that its top 1000 Redditors were active in and count the number of posts and comments published in each one [Appendix-4].
- For every Redditor of the 5000 we are focusing on, list the top 10 subreddits sorted by activity (number of posts and comments) [5].

Two data-cleaning actions were taken as a result of this analysis:

1. To reduce excessive noise from the data, we filtered out all the posts that had no textual body i.e., empty posts with titles only.
2. In addition, we decided to exclude subreddits that are directly or indirectly related to other musical genres and may potentially cause bias in our analysis. The full list of subreddits that were filtered out of the detailed users' activity can be found in 3].

We then updated the tables 6]7] and gave them a visual representation by plotting a bar chart per genre, that describes the number of authors that commented or posted at least once in the top non-musical subreddits. An additional stacked bar chart was plotted to highlight the similarities between the activity patterns of the different genres' fans. These visual representations exposed that the fans of all the genres share many of the same subreddits and demonstrated similar Reddit activity when excluding the musical-genre-related activity.

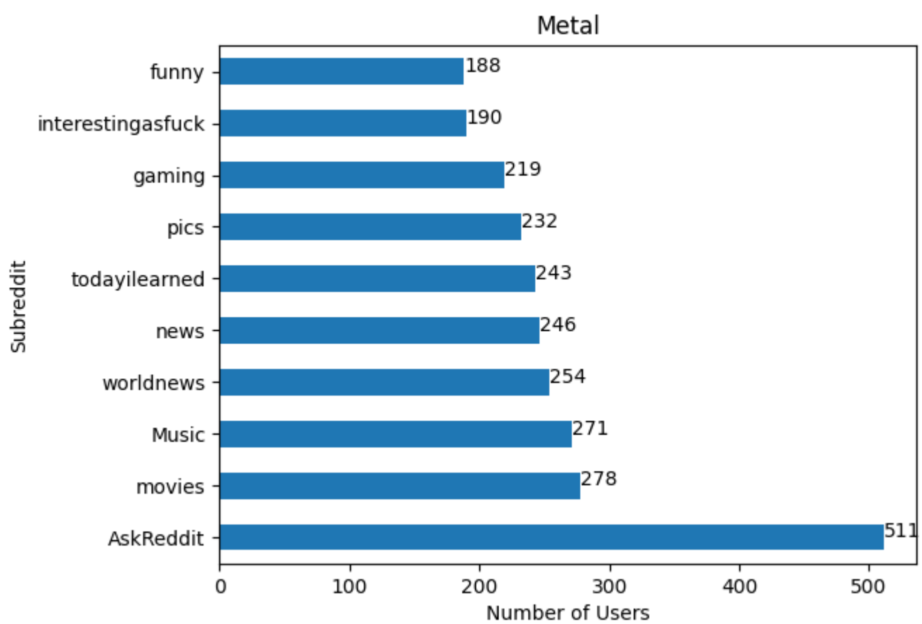

Figure 7: Number of authors in the top subreddits per musical genre community. This figure refers to the Metal fans community, similar distributions were demonstrated in all 5 genres and could be explored as part of the code 10]

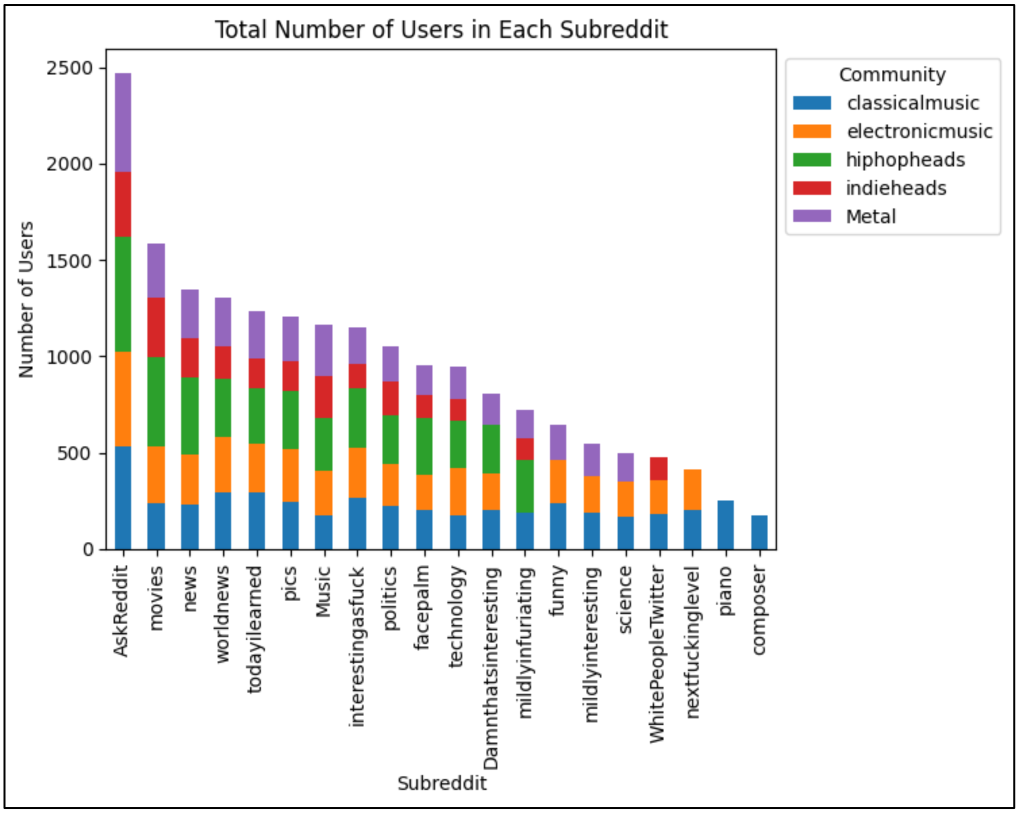

Figure 8: Top subreddits overlap: Stacked bar chart showing the number of authors that commented or posted at least once in the most popular subreddits, split by musical genre fans.

5.3 Dataset – Texts Written by Musical Genres Fans

The ideal dataset for our exploration would be a set of a few paragraph-length texts, written by a diverse population in terms of gender, demographic, age, and socio-economic status. The text will be tagged with a concrete and a single musical preference of the writer (and possibly for the Big Five personality traits scores, for sanity check). No writer belongs to more than one group of genres. The texts should not be related to music or the musical genres. While an effort was made to reach a data set at the abovementioned standard and minimize the noise and bias in the dataset, it is still expected to be unbalanced in terms of gender. The nature of data fetched from Reddit does not allow us to identify all the Redditors' genders without having them explicitly disclose it. Yet, it is known that 64% of Reddit users are male and only 36 % are female [53]. This distribution may have an impact on the research outcomes, as both genders tend to demonstrate different textual expression patterns. That being said, the effect is expected to be relatively minor on a large dataset like the one we gathered. To form a strong and stable baseline to start our work with, we will use existing tools and models that demonstrate good results in predicting personality traits for a given text, regardless of the writer's gender, age, or other demographic factors.

We revisited the freshly collected dataset, focusing on texts authored by highly active Reddit users within our selected specific musical genre communities. These users were exclusively active in their respective genre communities and not in other music-related subreddits. To enhance the dataset's quality and reduce the noise created by meaningless short texts, we filtered out texts shorter than 40 tokens and removed any duplicate entries. These filtering steps resulted in a more concise, yet higher quality dataset, described in Table 2.

	Classical	**Electronic**	**Hip-Hop**	**Indie**	**Metal**
Text Count after cleanup	170,251	97,063	121,538	84,314	102,650
Mean texts per user	173.37	109.55	122.39	96.8	107.94
Median texts per user	118.5	60.5	83.0	52.0	58.0

Table 2: Total number of posts and comments left after cleanup, filtering, and duplication removal, covering all 1K fans per genre, along with the mean and median number of texts per user for each genre.

6. Personality Traits Detection

Personality extraction is a multimodal problem, that can be approached by various methods. Input for personality extraction may vary from simple psychological surveys, through body language analysis [63], to eyeball movement [64]. Detecting personality traits from textual data has become an important area of research, especially with the advent of Natural Language Processing (NLP) techniques. The goal is to analyze written or spoken language to identify and predict various personality traits of individuals. These traits are often defined by psychological models like the Big Five personality traits, including Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. This section explores traditional and modern approaches to personality detection using NLP methods.

6.1. Traditional Approaches

Traditional approaches to personality trait detection have relied heavily on self-report questionnaires and psychometric tests. Instruments such as the Myers-Briggs Type Indicator (MBTI) and the Big Five Inventory (BFI) are commonly used to assess personality traits. While these methods are validated and widely used, they are often time-consuming, unscalable, and can be subject to biases, such as social desirability bias. In contrast, analyzing natural language data provides an unobtrusive and scalable way to infer personality traits.

6.2. Datasets in Use

Various datasets are used in NLP for personality detection. These datasets typically consist of textual data along with labeled personality traits (can be discrete or continuous), which are used to train and evaluate machine learning models. The following two datasets are the “gold standard” and are often used for models’ training. A high-level properties comparison of the two datasets is mentioned in 13.

6.2.1. Essays

One commonly used dataset is the Essays dataset, published in 1999 by Pennebaker and King [26]. It includes 2,468 essays or daily writing submissions, concluded by ~1.9 million words, collected between 1997 and 2004, and written by 2,468 students. These writing submissions were part of an ungraded psychology course requirement or assignment.

Students’ personality scores were determined by using the Big Five Inventory (BFI) [27], a 44-item self-report questionnaire that measures five personality traits. Each item consists of short phrases and is rated on a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

Each instance in the data set includes a filename or ID, the actual essay, and five classification labels representing the Big Five personality traits. The scores were converted to nominal classes with a median split [28], i.e. the final labels are either yes ('y') or no ('n') to indicate a high or low presence of a given trait – a binary label.

This dataset is considered a relatively rich source of data for training models to detect personality traits from text. However, it relies on BFI self-reports of a small number of individuals, and as described in 13, consists of relatively long texts with a mean length of 663.1 tokens.

While the Essays dataset provides a relatively rich source of data for training personality detection models, it has notable limitations. The small sample size of 2,468 students may limit generalizability, and the personality scores, based on BFI self-reports, can be biased. Converting scores to binary labels with a median split oversimplifies personality traits. The relatively long texts, averaging 663.1 tokens, require complex processing and can be computationally demanding. Additionally, the variability in the quality and effort of ungraded submissions may introduce noise into the dataset.

6.2.2. My Personality

Another significant piece of data is the MyPersonality dataset, which contains information from a Facebook application originally used by over 6 million users. David Stillwell and Michal Kosinski collected the data through this application, which implemented the Big Five test along with other psychological assessments [25]. User consent was obtained to record their data for research purposes.

Personality scores were assigned to users based on self-reports, assigned by a 100-item version of the IPIP personality questionnaire [65], along with several social network measures, including network size, density, brokerage, and transitivity [66]. The final dataset includes textual Facebook status updates, as well as continuous scores and Boolean labels for each personality trait per user. These Boolean labels were derived from the scores by a median split. This dataset is particularly valuable for research because it combines social media text with personality trait labels, allowing the study of personality in a more naturalistic context. Since the MyPersonality project was discontinued in 2018 [67], its data is no longer publicly available, and only a minor subset of 9,917 texts related to 250 users is available online.

Despite its value, the myPersonality dataset has several notable drawbacks. Firstly, personality scores are based on self-reported data, which can introduce biases and affect the accuracy of the results. Additionally, the publicly available subset of the dataset is limited to 250 users and fewer than 10,000 short texts, which significantly restricts the sample size, and the robustness of any analysis conducted. Furthermore, the personality scores serve as labels for users as single units rather than reflecting how each trait manifests in individual texts, potentially limiting the granularity and depth of personality trait analysis within the textual data.

7. Existing Tools for Personality Extraction from Text

As mentioned in the related work section, extensive research has been conducted on personality extraction from text. This body of work has led to the development of several publicly available "off-the-shelf tools" that can analyze text and predict the presence of the Big Five personality traits. These tools leverage advancements in computational methods and linguistic analysis to provide reliable and scalable personality assessments, making them valuable resources for both academic research and practical applications.

7.1. Tools Baseline Assessment

As our research does not focus on personality prediction from text, but on the correlation between musical preferences and personality as it is manifested in the texts we write, we wanted to extract personality traits from the Reddit dataset we collected, in the most accurate way we could. To do so, we surveyed the following existing personality extraction tools and evaluated them using the "gold standard" Essays and myPersonality datasets, mentioned above, for which we already have true labels for each of the Big Five traits per text. The results are listed in Table 3.

7.1.1. Psychology Insights

Tool #1 is an application created by Fugui Xing [68] for personality analysis, Big Five personality prediction, and emotion analysis. Powered by Azure Static Web App, Azure Function, React, and Machine Learning. This project hosted a full web application, including an API server and a React front end. We used the API server to predict the presence of the Big Five traits until the application was shut down, and with the guidance of the creator we used the trained model locally to complete our baseline work. As seen in the model-training repository [69] this tool was trained on the myPersonality dataset.

7.1.2. Personality Recognizer v1.03

The "*Personality Recognizer*" [70], published by Mairesse et al in 2007, is a Java command-line application that reads a set of text files and computes estimates of personality scores along the Big Five dimensions. It is based on Mairesse's work [71], which specifically mentions the Essays dataset as a used dataset. It has a live demo online, as shown in Figure 9, and it acts as a real black box, but was last updated in 2008, and showed poor results. **Therefore, we decided to exclude it from the baseline report.**

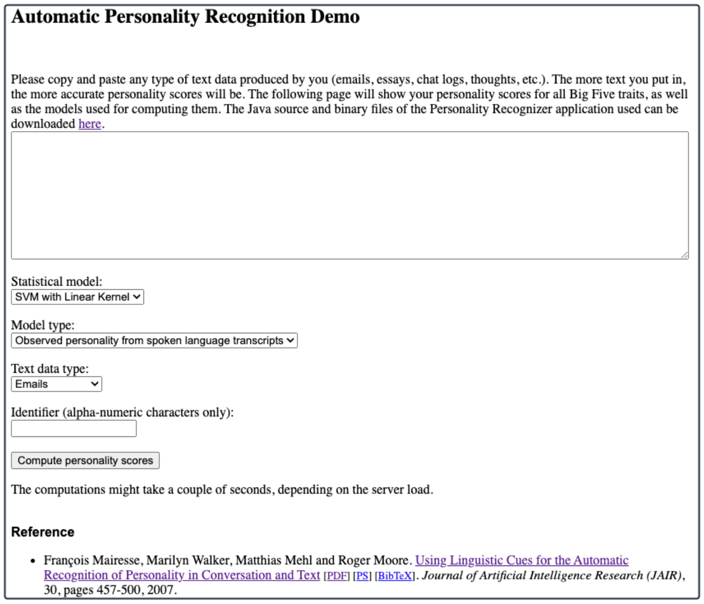

Figure 9: Personality Recognizer v1.03 online demo

7.1.3. Apply Magic Sauce

Apply Magic Sauce [72], marked as Tool #3, is a non-profit academic research project coordinated by the University of Cambridge Psychometrics Centre. The project exposes a live online platform that allows users to submit input from various sources like Facebook, X (formerly known as Twitter), LinkedIn, and open text, as shown in Figure 10, and get a full prediction of each of the Big Five personality traits and demographic conditions, demonstrated in Figure 11. The methods used were published by Kosinski and Stillwell [25] and claimed to be proven to beat human intuition. However, as declared on the application's website [72], its model is based on the myPersonality dataset.

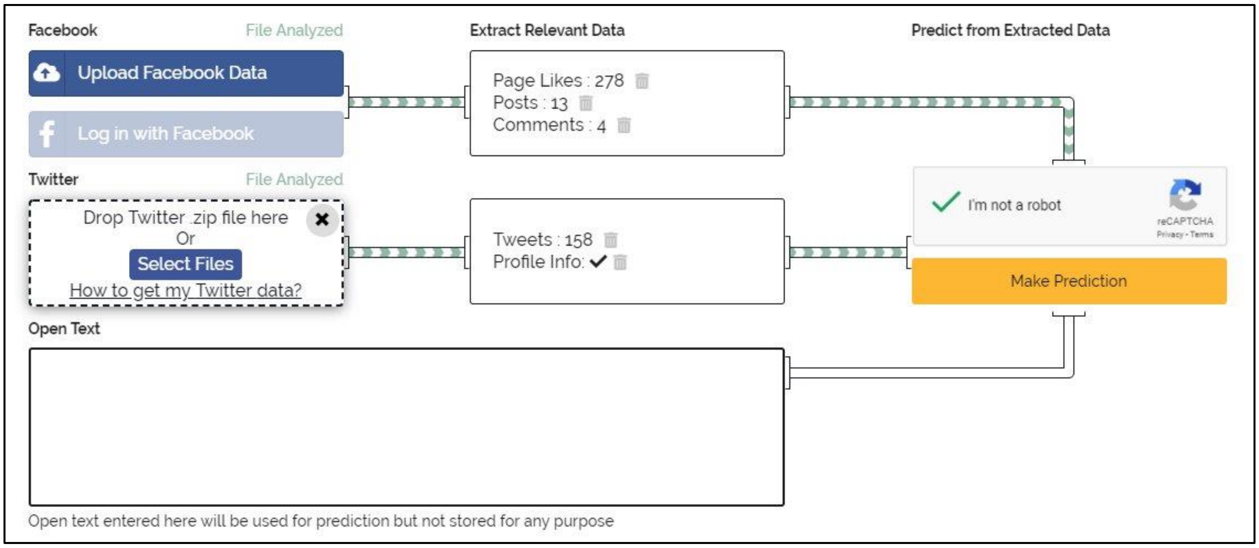

Figure 10: Apply Magic Sauce takes in textual input from various sources

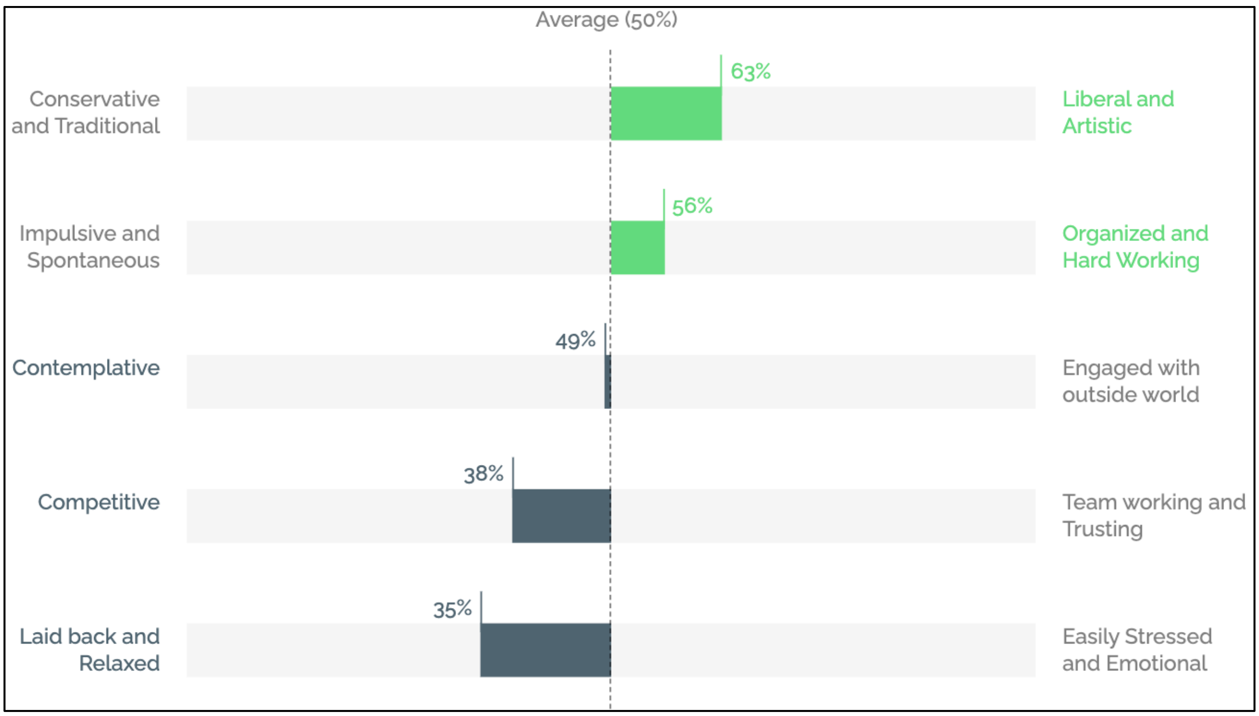

Figure 11: A sample of Apply Magic Sauce's Big Five personality prediction output

7.1.4. Jkwieser's Personality Prediction from Text

"*Personality Prediction from Text*" [73], marked as Tool #4, is an open-source project, inspired by the work of Majumder et al. [73] It was last maintained in 2020 and requires outdated libraries and hardware. Therefore, we had to employ it within a containerized environment, based on the publicly available docker image *jbei/scikit-learn:21.03* [75] to simulate the required conditions. It utilizes various NLP techniques such as SVM, Logistic Regression, BOW, Pre-Trained Glove, Random Forest, and more. It provides Boolean prediction for each of the Big Five traits (shown in Figure 12), and as declared in the project's git repository, one of the model's training sets was the Essays dataset.

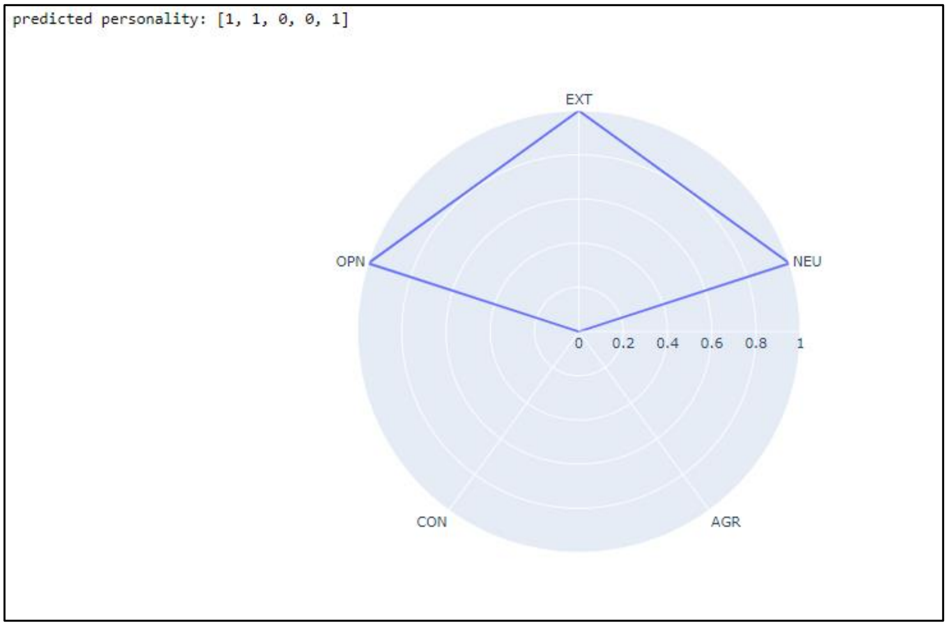

Figure 12: Example of a Boolean Big Five prediction by "Jkwieser's Personality Prediction from Text" as shown in the project's GitHub repository

7.2. Issues With Current Datasets

While several personality extraction tools are available online, it is evident that most of these tools have been at least partially trained using well-known datasets such as Essays or myPersonality. This reliance, coupled with the absence of high-quality, unseen datasets, renders baseline creation on these labeled datasets ineffective. Consequently, this introduces the new challenge of accurately measuring prediction accuracy on fresh, unseen datasets.

	Tool/Trait	EXT	NEU	AGR	CON	OPN
Essays	Tool #1	0.521	0.504	0.507	0.511	0.567
	Tool #3	0.495	0.549	0.510	0.538	0.557
	Tool #4	0.514	0.511	0.512	0.532	0.517
	Majority Vote	0.519	0.524	0.517	0.547	0.547
	*Logistic Regression**	0.55	0.59	0.55	0.51	0.62
MyPersonality	Tool #1	0.652	0.812	0.608	0.688	0.844
	Tool #3	0.596	0.56	0.568	0.572	0.544
	Tool #4	0.424	0.396	0.524	0.448	0.684
	Majority Vote	0.624	0.644	0.596	0.612	0.78

Table 3: Existing tools accuracy baseline; Tools marked with blue color signals that the tool trained on the tested dataset; The 'majority vote' is determined by the consensus of at least 2 of the 3 evaluated tools

8. Pivot: LLMs to the Rescue (5 classification problems)

Our initial efforts focused on evaluating existing tools for personality extraction from text, using gold-standard datasets such as Essays and myPersonality. However, we encountered significant limitations with these tools. All of them had been trained on the same datasets we intended to use for evaluation, which led to issues with overfitting and biased performance metrics. Additionally, the results from these evaluations were only moderate, failing to meet our expectations for accuracy and reliability.

Given these challenges, we decided to pivot our approach and leverage the capabilities of large language models (LLMs) like GPT [76], Bard (Gemini) [77], and Bing-CoPilot [78]. The decision was motivated by the following factors:

1. State-of-the-Art Performance: LLMs have demonstrated superior performance in a variety of natural language processing tasks, including text generation and understanding. Recent research by Peters and Matz (2024) [37] was among the first to apply GPT-3.5 and GPT-4 to predict the Big Five personality traits from users' Facebook status updates with a somewhat weak correlation of $r = 0.29$ between LLM-inferred and self-reported trait scores. Although this is slightly lower than supervised models (e.g., Park et al. [36] with $r = 0.37$), it underscores the potential of LLMs in generating personality-correlated text.
2. Scalability and Adaptability: Unlike traditional models, LLMs can be easily adapted to new tasks and datasets without the need for extensive retraining, offering a scalable solution for generating varied and high-quality text data. Safdari et al. (2023) [38] demonstrated the reliability and validity of personality measurements in LLM outputs under specific prompting configurations, emphasizing the flexibility of LLMs in generating diverse textual content.
3. Rich Textual Content Generation: LLMs can generate diverse and nuanced textual content, which can be used to create new datasets that exhibit strong correlations with specific personality traits. The work by Safdari et al. [38] also explored shaping personality traits within LLM outputs, highlighting the potential for generating texts that mimic specific personality profiles.

To harness the potential of LLMs, we devised a systematic approach to generate textual content that could be used to train and evaluate our models for each of the Big Five personality traits. Here are the steps we undertook, as illustrated in Figure 13:

1. Definition Compilation: We gathered detailed definitions for each of the Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) from academic and online resources.
2. Readable Definition Construction: These definitions were manually concatenated into a readable and comprehensive format to serve as the basis for our prompts.
3. Prompt Creation: For each personality trait, we wrote a primary prompt that started with the trait's definition and ended with a request for the LLM to generate ten paragraphs

demonstrating a high score for the trait. Similar prompts were created to generate texts with low trait scores.

4. Iterative Refinement: The prompts were refined iteratively. We modified the requests to ensure the generated texts were unique, random, and lengthy and avoided repetition. We also ensured that the texts were written from diverse perspectives and character profiles.

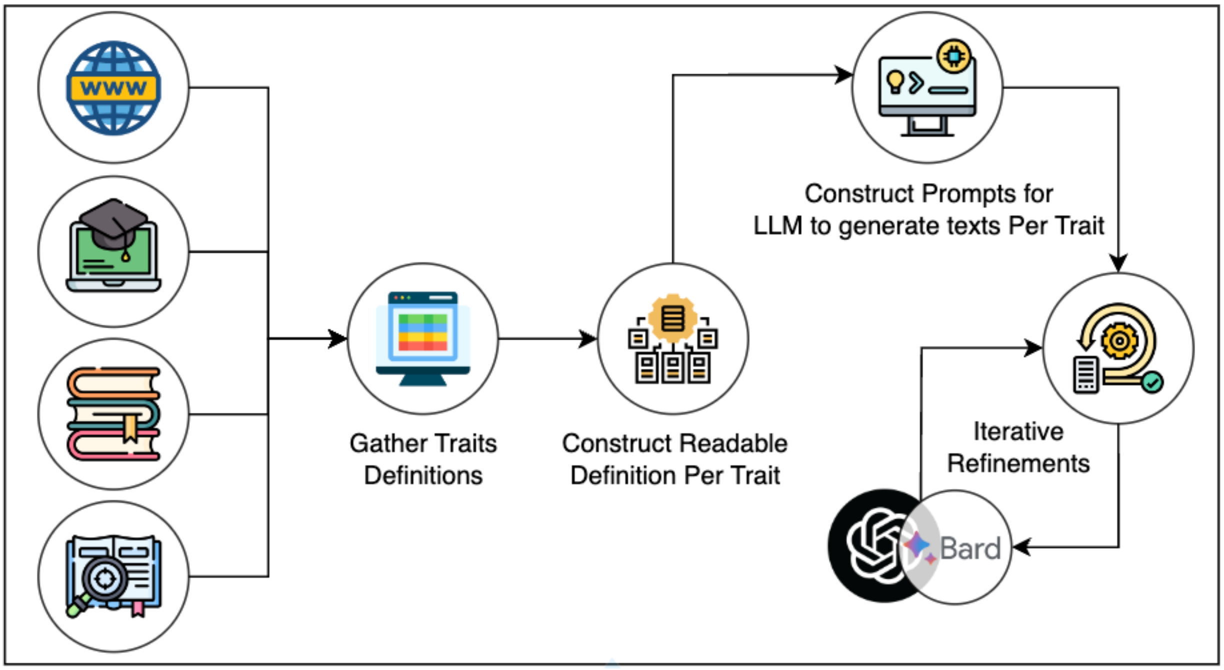

Figure 13: LLM-Based Dataset Generation Flow

We employed three popular free-to-use LLMs for our text-generation tasks:

- OpenAI's Chat-GPT3.5 [76]: Known for its conversational capabilities and contextual understanding.
- Google's Bard (Gemini) [77]: Renowned for its advanced language comprehension and generation.
- Microsoft's Bing-CoPilot [78]: Integrated with search capabilities to enhance information retrieval and text generation. After a few experiments, we decided to neglect the usage of Bing-CoPilot as it demonstrated inferior results, compared to the other models.

By pivoting to LLMs, we aimed to create a rich and varied dataset that would enable us to assess the existing tools we tried to use for baseline creation and train new models for personality trait extraction. This approach allowed us to overcome the limitations of the existing tools and leverage the strengths of state-of-the-art LLMs to address our five concrete classification problems.

8.1. The Datasets Generation

The dataset generation process for this study involved multiple steps to ensure the creation of high-quality and diverse texts that reflect various personality traits as per the Big Five personality model. Below is a detailed explanation of the process:

8.1.1. Gathering Definitions

We began by gathering definitions for the Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) from both academic and online resources. These definitions were compiled into an Excel table, categorizing them by source, trait, general description, and characteristics of high and low presence of each trait. This provided a comprehensive understanding of each trait and served as the foundation for crafting the prompts used in the text generation. A subset of the table is shown in Table 4, and the full set of definitions is available in [11].

Trait	General Description	High	Low
OPN	Openness (also referred to as openness to experience) emphasizes imagination and insight the most out of all five personality traits. People who are high in openness tend to have a broad range of interests. They are curious about the world and other people and are eager to learn new things and enjoy new experiences. People who are high in this personality trait also tend to be more adventurous and creative. Conversely, people low in this personality trait are often much more traditional and may struggle with abstract thinking.	Very creative	Dislikes change
		Open to trying new things	Does not enjoy new things
		Focused on tackling new challenges	Resists new ideas
		Happy to think about abstract concepts	Not very imaginative
			Dislikes abstract or theoretical concepts
CON	Among each of the personality traits, conscientiousness is one defined by high levels of thoughtfulness, good impulse control, and goal-directed behaviors. Power RA, Pluess M. Heritability estimates of the Big Five personality traits based on common genetic variants. Highly conscientious people tend to be organized and mindful of details. They plan ahead, think about how their behavior affects others, and are mindful of deadlines. Someone scoring lower in this primary personality trait is less structured and less organized. They may procrastinate to get things done, sometimes missing deadlines completely.	Spends time preparing	Dislike's structure and schedules
		Finishes important tasks right away	Makes messes and doesn't take care of things
		Pays attention to detail	Fails to return things or put them back where they belong
		Enjoys having a set schedule	Procrastinates important tasks
			Fails to complete necessary or assigned tasks
EXT	Extraversion (or extroversion) is a personality trait characterized by excitability, sociability,	Enjoys being the center of attention	Prefers solitude

Trait	General Description	High	Low
	talkativeness, assertiveness, and high amounts of emotional expressiveness.1 People high in extraversion are outgoing and tend to gain energy in social situations. Being around others helps them feel energized and excited. People who are low in this personality trait or introverted tend to be more reserved. They have less energy to expend in social settings and social events can feel draining. Introverts often require a period of solitude and quiet in order to "recharge."	Likes to start conversations	Feels exhausted when having to socialize a lot
		Enjoys meeting new people	Finds it difficult to start conversations
		Has a wide social circle of friends and acquaintances	Dislikes making small talk
		Finds it easy to make new friends	Carefully thinks things through before speaking
		Feels energized when around other people	Dislikes being the center of attention
		Say things before thinking about them	
AGR	This personality trait includes attributes such as trust, altruism, kindness, affection, and other prosocial behaviors.1 People who are high in agreeableness tend to be more cooperative while those low in this personality trait tend to be more competitive and sometimes even manipulative.	Has a great deal of interest in other people	Takes little interest in others
		Cares about others	Doesn't care about how other people feel
		Feels empathy and concern for other people	Has little interest in other people's problems
		Enjoys helping and contributing to the happiness of other people	Insults and belittles others
		Assists others who are in need of help	Manipulates others to get what they want
NEU	Neuroticism is a personality trait characterized by sadness, moodiness, and emotional instability. Individuals who are high in neuroticism tend to experience mood swings, anxiety, irritability, and sadness. Those low in this personality trait tend to be more stable and emotionally resilient.	Experiences a lot of stress	Emotionally stable
		Worries about many different things	Deals well with stress
		Gets upset easily	Rarely feels sad or depressed
		Experiences dramatic shifts in mood	Doesn't worry much
		Feels anxious	Is very relaxed
		Struggles to bounce back after stressful events	

Table 4: Personality Traits Definitions

8.1.2. Crafting Prompts

Recognizing that personality prediction from text can be viewed as five distinct Boolean classification problems (one for each trait: high or low presence), we created sets of high presence and low presence texts for each trait. For example, for Extraversion, we generated texts “written by” people who are highly extroverted and separated texts “by” those who rank low on extroversion, i.e. considered introverts.

Based on the gathered definitions, we crafted a main prompt for each trait, defining what the trait means and how a person with a high or low presence of the trait behaves. These prompts varied in length from 391 to 488 tokens. For example, the prompt for openness included:

> *"Openness (also referred to as openness to experience) emphasizes imagination and insight the most out of all five personality traits. People who are high in openness tend to have a broad range of interests. They are curious about the world and other people and are eager to learn new things and enjoy new experiences. People who are high in this personality trait also tend to be more adventurous and creative. Conversely, people low in this personality trait are often much more traditional and may struggle with abstract thinking..."*

The prompts were designed to ensure the generated texts reflected the trait without focusing explicitly on it. This approach helped in producing diverse and authentic representations of a high and low presence of each trait. To achieve this, we iterated over various personas and contexts. For example, for high openness, we requested:

> *"I need to collect 10 paragraphs of 70-150 words each, "written by" people with high openness personality. All the paragraphs should be very diverse and should not be repeated at all. If possible, don't even repeat sentences. Let's start with the first paragraphs and then continue with 9 more iterations."*

The five main prompts are available in 12].

8.1.3. Prompts Application

Using the crafted five main prompts, we used Chat-GPT 3.5 and Bard via their publicly available web user interfaces and applied each prompt individually. By repetitive manual iterations, we improved the prompt until we achieved satisfactory results, based on manual inspection of 30 samples out of more than 100 texts generated for each trait. We ensured the texts were diverse enough with minimal structural, or linguistic repetition.

To ensure the data generated was not overly focused on the trait but instead reflected it naturally, we continuously refined our prompts. We asked the large language models to write as if they were different personas, iterating over a variety of topics and writer profiles.

For instance:

1. Texts written by High school teenagers
2. Texts written by Parents from various socioeconomic backgrounds
3. Texts written by Sports lovers (fans or practitioners)
4. Texts written by Tech-savvy individuals

5. Texts written by Politicians
6. Texts written by Freelancers
7. Texts about global warming
8. Texts about History
9. Texts about Finance
10. Texts about News
11. Texts about Art
12. Texts about Sound system
13. Texts about Science fiction
14. Texts about Rare animal

This process allowed us to generate a rich and varied dataset, ensuring the texts were reflective of different personas without being repetitive or predictable. The generated texts, sampled in 14, were then reviewed to confirm they met the criteria for each personality trait, and some necessary adjustments, like similar text filtering, were made to maintain diversity and authenticity.

8.2. Evaluation: Existing Tools and LLM Datasets

To evaluate the dataset's quality and the tools discussed in Section 7, we began by focusing on a single personality trait: Openness to Experience (OPN), and the approach's overview is described in Figure 14. Text embeddings, which are vector representations of text, have been widely proven effective in numerous Natural Language Processing (NLP) tasks, including semantic search, text classification, and similarity detection. By converting text into dense vector representations, embeddings capture the underlying semantics, enabling machines to understand and process human language more effectively. This effectiveness is well-documented in various studies that demonstrate the ability of text embeddings to improve the performance of NLP models across a wide range of tasks [79][80].

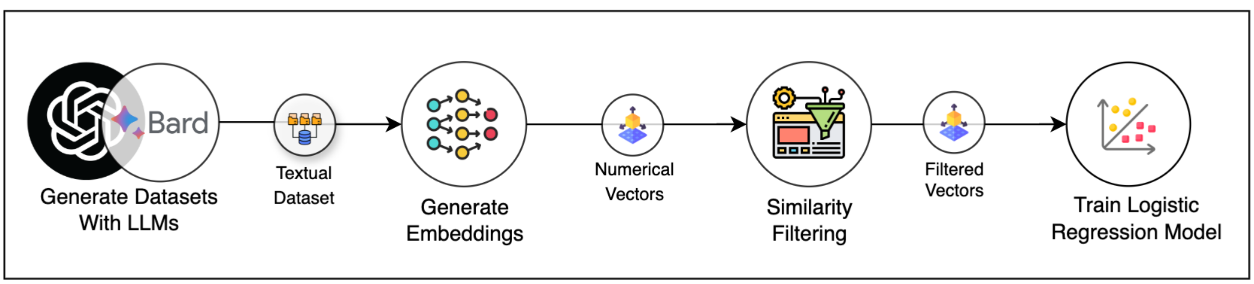

Figure 14: Workflow Overview for Generating Our Classification Models (High vs. Low Presence of a Single Personality Trait)

To generate embeddings for our LLM-generated dataset, we used the sentence-transformers Python package [56], a robust tool designed for generating high-quality text embeddings, combined with the *intfloat/e5-large-v2* model. This text embedding model was curated by

Microsoft's research team and is based on weakly supervised contrastive pre-training and features 24 layers with 330M parameters [57]. We specifically chose the *intfloat/e5-large-v2* model due to its exceptional performance in a variety of text embedding tasks. It has outperformed many leading models like BERT, Snowflake's Arctic, and Meta's DRAGON+ model [81] in benchmarks such as the MTEB [82] and BEIR [83], and it not only substantially outperforms existing models with similar sizes, but also match the results of much larger models, indicating its superior ability to capture semantic nuances and provide reliable, high-quality embeddings [57].

The *intfloat/e5-large-v2* model's embeddings are not only dense and high-dimensional but also designed to better preserve semantic relationships between texts, making it particularly well-suited for our task of predicting personality traits from the language used by different musical genre fans. Additionally, the model's efficiency in handling large datasets made it a practical choice for our project, allowing us to embed extensive amounts of text data without sacrificing performance.

To ensure the quality of our datasets, we calculated the cosine similarity between each pair of text embeddings within each dataset and removed any text pairs with a similarity greater than 0.95, thereby reducing redundancy and the possibility of overtraining or bias models. Cosine similarity measures the similarity between two non-zero vectors, calculating the cosine of the angle between them, ranging from -1 to 1, where 1 means they are identical and 0 means no similarity.

We then proceeded to train logistic regression models to classify texts as either high or low in openness, leveraging the filtered, semantically rich embeddings generated by the *intfloat/e5-large-v2* model. This approach allowed us to extract meaningful insights from the text data, grounded in a robust and well-tested methodological framework.

Logistic regression was chosen as the classifier for this task due to its simplicity, efficiency, and interpretability. Logistic regression is a linear model that is well-suited for binary classification problems, which aligns with our goal of classifying texts as either high or low in specific personality traits. Despite being a relatively straightforward model, logistic regression offers several advantages: it is computationally efficient, making it ideal for large-scale datasets, and it provides interpretable results, allowing us to easily understand the relationship between the input features (text embeddings) and the predicted outcomes (personality traits). This interpretability is crucial in our research, as it helps validate the results and provides insights into the factors contributing to personality trait predictions.

We trained logistic regression models on the datasets generated by the language models, containing texts with high and low OPN for each model. As shown in Table 56 the model trained on the Chat-GPT-generated data outperformed the model based on Bard data, with an accuracy of 0.847, when tested on the unseen test set generated by the alternative model. Despite both datasets being generated by LLMs, no significant suspicious similarity was identified between them that could account for the superior performance of the GPT-based logistic regression model.

Train\Test	Chat-GPT	Bard
Chat-GPT	0.98	0.847
Bard	0.50	0.827

Table 56 Logistic regression models' accuracy when trained and tested on LLM-generated OPN datasets

To identify the most accurate tool or model among those explored in Section 7 and the logistic regression model trained on the Chat-GPT dataset (referred to as L.R. **A.**), we evaluated their performance in predicting high and low OPN on the unseen test set generated by Bard. For robustness, we also tested a logistic regression model trained on the Chat-GPT dataset combined with the myPersonality dataset (referred to as L.R. **B.**).

L.R. **A.** achieved an overall accuracy of 0.847, with 100% precision and recall when predicting Bard's High-OPN set (51 out of 51 samples). For Bard's Low-OPN set, it maintained 100% precision with a recall of 0.763, correctly identifying 71 out of 93 samples. In contrast, L.R. **B.** performed significantly worse, with an accuracy of 0.526. Tool #1, described in Section 7.1.1, scored 0.34, while Tool #3 (Section 7.1.3) performed slightly better at 0.43 accuracy. Tool #4 (Section 7.1.4) had the lowest accuracy at 0.32, as detailed in Table 6.

Rank	Model	Accuracy
1	L.R. **A.**	0.847
2	L.R. **B.**	0.526
3	Tool #3	0.452
4	Tool #1	0.347
5	Tool #4	0.320

Table 6: models' and tools accuracy, predicting high and low OPN on Bard-generated unseen OPN dataset.

Due to the superior performance of **L.R. A.**, trained on the Chat-GPT dataset, we chose it for further extraction of personality traits from our Reddit dataset, as discussed in Section 5. As mentioned, our evaluations have focused on the OPN personality trait, with less emphasis on the other four traits: Conscientiousness, Extraversion, Agreeableness, and Neuroticism. To fully leverage the capabilities of our chosen tool, **L.R. A.**, before going back to our Reddit dataset, we developed and trained specific models for each of these traits. By training on datasets generated by Chat-GPT and testing on unseen datasets produced by Bard, we obtained excellent accuracy results, as detailed in Table 7.

Trained on \ Trait	EXT	OPN	NEU	AGR	CON
(L.R. A.) GPT	0.963	0.874	0.991	0.959	0.934
(L.R. B.) GPT + myPersonality	0.848	0.526	0.731	0.939	0.954

Table 7: Logistic Regression models per personality traits accuracy predicting the Big Five personality traits on the unseen Bard-generated test set

In conclusion, we evaluated several tools for predicting personality traits from text. By generating textual content strongly associated with high and low levels of openness to experience (OPN), we trained logistic regression models to classify texts by OPN levels and compared their performance with existing tools. The best performing of all we chose to proceed with is the logistic regression model trained on the dataset generated by Chat-GPT.

9. Back to the Reddit Dataset

Using the same sentence-transformer python library [56] with the intfloat/e5-large-v2 model, we vectorized the filtered Reddit dataset which its collection is detailed in section 5, and each user got multiple vectors, one per comment or post.

Index	Community	Username	Text	Origin Subreddit	# of long texts
0	classicalmusic	Radaxen	I remember doing this with WC3 Dota 1 Lifestea...	DotA2	339
170251	electronicmusic	headphase	This Monday, June 12, r/BedStuy will become un...	BedStuy	305
267314	hiphopheads	thanks_bruh	Whats good y'all. I'm looking to make friends ...	washingtondc	50
388852	indieheads	simco1974	2008 or so, I was digging around at a local re...	TheeOhSees	1
473166	Metal	sam1oq	I found out that apparently I was still paying...	LifeProTips	309

Table 8: Data frame sample, including a unique embedding vector per text. The column # of long texts refers to the number of texts by the user that has at least 40 tokens.

The next step was to utilize the best-performing personality extraction tool we evaluated: **L.R. A.**, trained on datasets generated by Chat-GPT. We developed five distinct **L.R. A.** models, each trained on one of the Big Five personality traits datasets. Although these models are trained as binary classifiers, they produce a posterior probability indicating the likelihood of an observation belonging to the positive class. This characteristic makes the models particularly convenient for extracting a continuous (rather than binary) score for each trait. We had the models predict the probability of each text displaying a high presence of its corresponding trait. Subsequently, for each user, we set the probability of scoring high on each trait by calculating the mean probability value across all their texts. A sample of these results is shown in Table 9, where we can already see some of the differences detailed in the results section, for instance, Metal lovers' likelihood to manifest higher levels of neuroticism than fans of other genres.

community	user	texts	OPN	EXT	NEU	AGR	CON
classicalmusic	Radaxen	339	0.42	0.47	0.51	0.32	0.33
electronicmusic	headphase	305	0.41	0.52	0.51	0.26	0.30
hiphopheads	acasovoycayendo	145	0.42	0.53	0.50	0.29	0.34
indieheads	GeorgeTaylorG	12	0.44	0.48	0.53	0.31	0.31
metal	sam1oq	309	0.40	0.46	0.55	0.27	0.31

Table 9: Data frame sample showing mean probability for each trait (Openness, Extraversion, Neuroticism, Agreeableness, and Conscientiousness) per user. For instance, we can see the user "sam1oq", a Metal fan, exhibits the highest likelihood of neuroticism compared to fans of other genres.

10. Results

10.1. Overview

In this study, our objective is to explore the relationship between musical preferences and personality traits rather than focusing on individual users. To do so, we utilized the logistic regression models we developed (L.R. A.), which outperformed the alternative tools evaluated. We predicted a score for each of the Big Five personality traits for every text. Table 10 shows examples of concrete texts that scored high or low for the presence of each personality trait. We then averaged the prediction scores across all texts authored by a given user. This approach allowed us to assign each user a personality trait score. By aggregating these user-level scores by musical genre we were able to analyze personality trait probability predictions at the community level, revealing several significant differences among fans of different musical genres.

Trait & Level	Sample Text	Score
HIGH OPN	Wow....there is so much awesome in this artwork. Pretty amazing GJJvO could render such distinct subjects as the numerous blooming flowers, the Greek vase fragmented story and the bronze relief sculpture. Mind blowing visually. I bet there's language symbolized in the flowers interacting with the decorative elements.	0.754
LOW OPN	Okay but we're not talking about a person who doesn't care, but just doesn't want to identify as left right or center as our society defines them. Change vs Stability also doesn't really make sense. The right in America wants things to change, the left also wants things to change, even the center wants things to change. They just don't agree on what the change should be and what goals the change should accomplish.	0.189
HIGH EXT	No game will ever use the full potential of a new system in the first year, but what we're seeing is already a huge improvement over last-gen. Not only are these games running at higher resolutions with significantly higher detail, but the vast majority are also running at double the frame rate. This is a really big deal.	0.802
LOW EXT	I've felt reprehension about going to a park, sitting on the grass and meditating. Precisely because of this reprehension, I decided to do it, and to make that feeling the object of my meditation. I went and did it, and obviously no one cared, no one probably noticed, everyone's minding their own business and I sat near a tree and did like 10 minutes closed eyes meditation. When I heard people walk near I felt shame rising and the urge to pretend to be busy doing something else, but i just let the footsteps walk away and again, nothing happened. Afterwards I just got up and walked home.	0.152
HIGH NEU	feel like i'm going crazy. everything is up in the air. no idea where i'll be in august. i'm applying for jobs and my soul tears apart each time because the process is designed to be fucking excruciating and make you feel worthless [so they can hire you cheaper]. similarly sent in my masters application. it is so awful i hate it	0.811
LOW NEU	This is great. Eve and Amanda and Elza Brabant talk about the 3 reasons we eat: we eat for nutrients, we eat for pleasure, and we eat for community. I like how you are thinking: I think what you are doing enhances your eating. Best.	0.184
HIGH AGR	The more I look at this landscape, the more in revealed in its shaded depths and sunny ridge flora. The range of autumn colors allow a wide range of lighter colored grass blades and plant stems that help direct my viewing pathways.	0.743

Trait & Level	Sample Text	Score
LOW AGR	IMO there was a mix of them being the kind of developer that goes by "if it works, who cares about the code" and being totally in the known that writing shitty code would made them the owners of whatever crap they wrote. All their code was absolutely spaghetti and any trace of unit/integration test were left to the rest of us.	0.106
HIGH CON	Alan Walker's incredibly detailed and insightful 3-part biography of Liszt, as well as his works on Chopin and others, provide a musical bridge from the 1830s right up to the present day, via Walker's remarks and the recordings of those who studied under Liszt. One can learn so much about how to draw true music out of piano works and one's own life experiences from Walker's views. Through his generous lessons over a long life, Liszt left a mighty legacy of pianistic and musical beauty that will live on forever.	0.737
LOW CON	Yeah, it definitely sucked. I remember a girl I was dating taking me there, and looking around at how empty it was on a Saturday, I just said "this isn't going to last." She got mad at me, but that was six months before the closure. Very cool idea, I just think it was in the wrong decade for it to exist. I never considered looking at the state of the cars--that'll be interesting to look at next time!	0.132

Table 10: Example texts illustrating varying levels of presence for each personality trait (Openness, Extraversion, Neuroticism, Agreeableness, and Conscientiousness), accompanied by a probability score indicating the likelihood of high trait presence. The scores were assigned using our Logistic Regression Analysis (L.R. A.) model.

Some of these findings align with common intuitions: for example, Classical music enthusiasts tend to exhibit the highest levels of agreeableness, while Hip-Hop fans display the lowest levels of agreeableness compared to all other genres. Additionally, Classical music enthusiasts generally show lower levels of extroversion, and Metal fans tend to have higher levels of neuroticism. These are just a few of the findings, with more detailed results comprehensively summarized in Table 11, which outlines the correlations between musical genres and personality traits. The numbers in the table represent mean and median values over user scores grouped by genre.

Genre	Measure	OPN	EXT	NEU	AGR	CON
Metal	mean	0.437	0.508	0.518	0.297	0.313
	median	0.433	0.510	0.519	0.294	0.309
Classical	mean	0.444	0.480	0.502	0.317	0.330
	median	0.439	0.480	0.503	0.312	0.325
Electronic	mean	0.443	0.524	0.509	0.300	0.311
	median	0.435	0.523	0.509	0.294	0.306
Hip-Hop	mean	0.418	0.522	0.520	0.277	0.304
	median	0.414	0.523	0.520	0.274	0.301
Indie	mean	0.439	0.511	0.515	0.302	0.310
	median	0.434	0.513	0.516	0.297	0.306

Table 11: Probability to score high on a personality trait, out of the big five traits (Openness, Extraversion, Neuroticism, Agreeableness, and Conscientiousness) for each musical genre.

To further illustrate these trends, Figure 15 and Figure 16 present box plots that graphically highlight these intuitions, providing a clear visual representation of the variations in personality traits across different music genres.

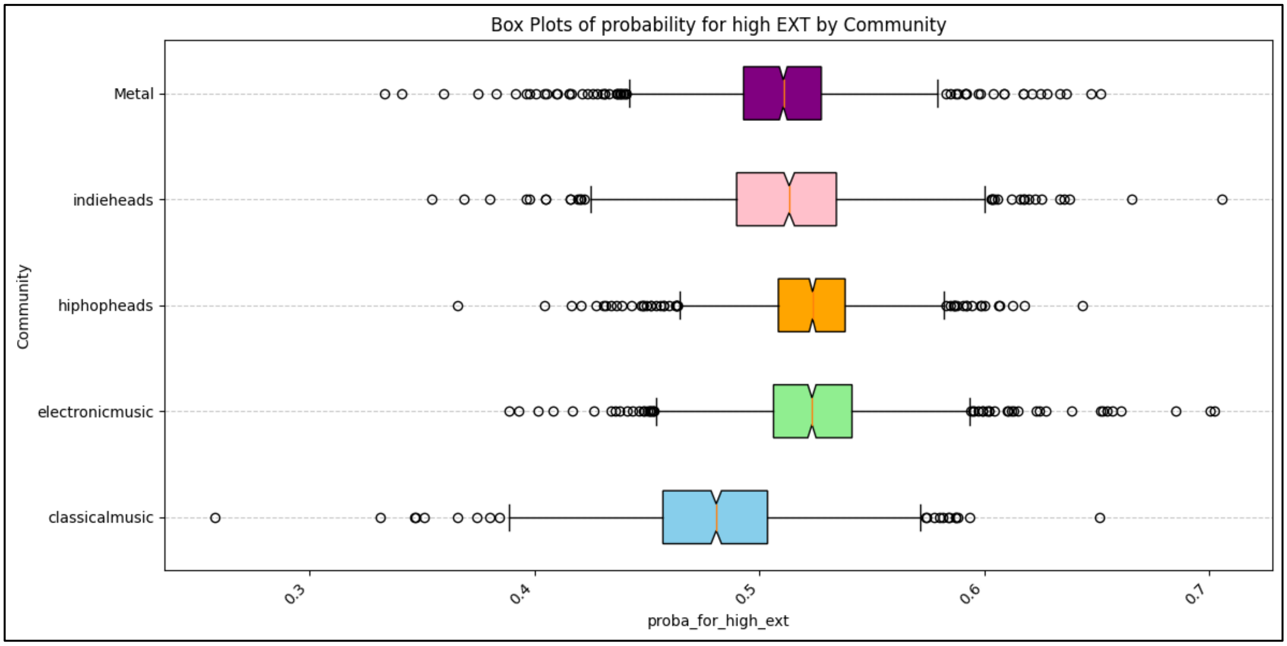

Figure 15: Extraversion by genre box plot, showing how Classical music fans demonstrate less extroversion-related behaviors in their texts

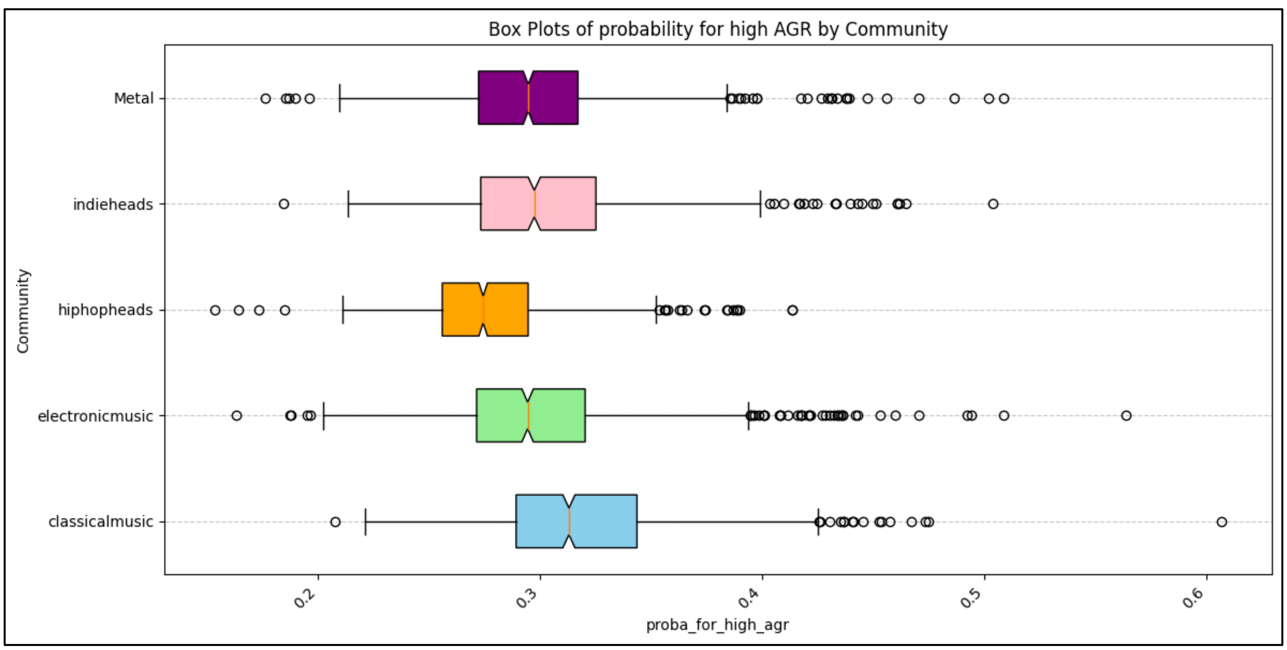

Figure 16: Agreeableness by genre box plot, showing how Hip-Hop fans tend to be less agreeable than the others while Classical fans are the opposite.

10.2. Statistical Tests: ANOVA

With nearly 1,000 probability predictions for each trait per musical genre, we aim to evaluate the significance of differences between the groups' scores. To achieve this, we will employ the Analysis of Variance (ANOVA) test. ANOVA is a statistical method that compares means across multiple groups to identify statistically significant differences. It analyzes both within-group and between-group variation to determine if observed differences are due to actual effects or random variation. Particularly useful for experiments involving three or more groups, ANOVA extends the capabilities of the two-group t-test. The primary output, the F-statistic, indicates whether the means of the groups are significantly different. A p-value less than 0.05 typically indicates statistical significance.

Personality Trait	F-Value	P-Value
OPN	82.73	5.73e-68
EXT	258.38	7.18e-201
NEU	55.74	5.47e-46
AGR	129.72	3.30e-105
CON	109.00	6.03e-89

Table 12: ANOVA test results

The ANOVA test results detailed in Table 12, strongly suggest that there are significant differences in the means across the groups for each of the Big Five personality traits. The extremely low P-values confirm that these differences are highly statistically significant. The high F-values, especially for traits like EXT and AGR, highlight that these traits vary more distinctly between the groups under analysis.

Genre	Sample Size
Metal	951
Classical	982
Electronic	886
HipHop	993
Indie	871

Table 13: Final sample size per musical genre

10.3. Statistical Tests: T-TEST and Cohen's-D Effect Size

While ANOVA provides a broad analysis by testing whether there are any significant differences across multiple groups, it does not specify where those differences lie. To gain a more detailed understanding, we also performed T-tests, which allowed us to compare the means of pairs of groups directly. This focused analysis helps determine not only whether the means differ between two groups but also in which direction the difference lies. By zooming in on these pairwise comparisons, we can better interpret the specific nature of the differences observed. Additionally, we calculated Cohen's d to quantify the effect size, providing insight into the magnitude of these differences.

The t-test is a statistical method used to determine if there is a significant difference between the means of two groups. It calculates a t-statistic, which reflects the magnitude of the difference relative to the variability within the groups. A high absolute t-statistic value indicates a stronger difference between the groups, and the associated p-value shows the probability that this difference is due to random chance.

Cohen's d is a measure of effect size that quantifies the magnitude of the difference between two groups in terms of standard deviations. It helps to understand the practical significance of the observed differences. The test's formula is described in Equation 1.

$$d = \frac{M_1 - M_2}{SD_{pooled}} \qquad \text{Where} \qquad SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

Equation 1: Cohen's D test formula

M_1 and M_2 stand for the means of the groups and the SD_{Pooled} is the pooled standard deviation. SD_1 and SD_2 stand for the standard deviation of the two groups, and n_1 and n_2 are the sample sizes of the groups. Cohen's d values are interpreted as follows: a small effect of around 0.2, a medium effect of around 0.5, and a large effect of 0.8 or higher. This test is useful in determining the practical significance of research findings. A larger Cohen's d value indicates a larger difference between the groups, relative to the variability within the groups.

In addition to the ANOVA test, we used both the t-test and Cohen's d to assess the statistical and practical significance of the differences between predictions for each musical genre. The t-test results provide insight into whether the differences are statistically significant, while Cohen's D helps gauge the practical importance of these differences. Figure 17 describes the T-Test results via a combined heatmap, in which darker cells indicate a higher degree of significance of the results. The cell's color is affected by the t-statistics values, and the p-values are shown in parentheses. The t-statistics sign indicates the direction of difference, so, for example, `f-statistics=-31.82`, for Extroversion in Classical vs HipHop, means that Classical is less likely to demonstrate the trait.

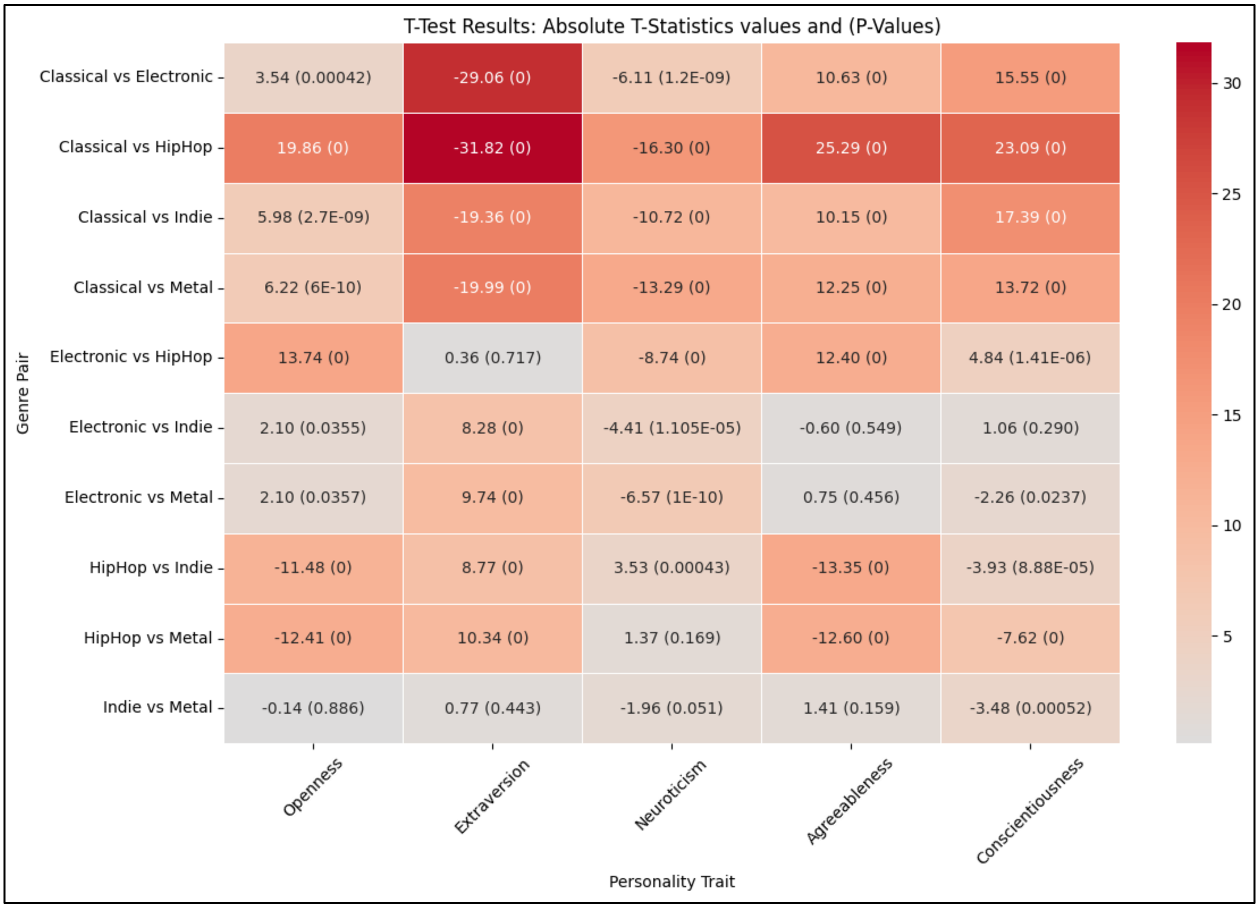

Figure 17: T-Test results heatmap

The Cohen's D test results exhibit a strong correlation with the T-Test findings, highlighting the most substantial effect sizes between classical and hip-hop genres concerning Extroversion, Agreeableness, and Conscientiousness traits. Notably, a significant effect was observed between classical and electronic music fans with respect to Extroversion. Figure 18 illustrates various genre combinations associated with large effect sizes. The d-value sign indicates the direction of difference, so for example $d=-1.508189055$, for Extroversion in Classical vs HipHop, means that the mean EXT score for the Classical music group is lower than the mean EXT score for the HipHop group. In other words, individuals in the Classical music group tend to score lower on EXT compared to those in the hip-hop group.

The analysis of Cohen's D test results reveals significant associations between musical preferences and personality traits. Classical listeners are notably less extroverted compared to both hip-hop and electronic listeners, as indicated by large negative effect sizes. Additionally, Classical listeners exhibit higher levels of agreeableness, particularly when contrasted with hip-hop listeners. In terms of openness, hip-hop shows the lowest presence compared to the other genres, especially Classical and Electronic. Meanwhile, hip-hop and Metal listeners display higher levels of neuroticism when compared to Classical listeners. These findings suggest that the presence or absence of certain personality traits is strongly influenced by musical preferences.

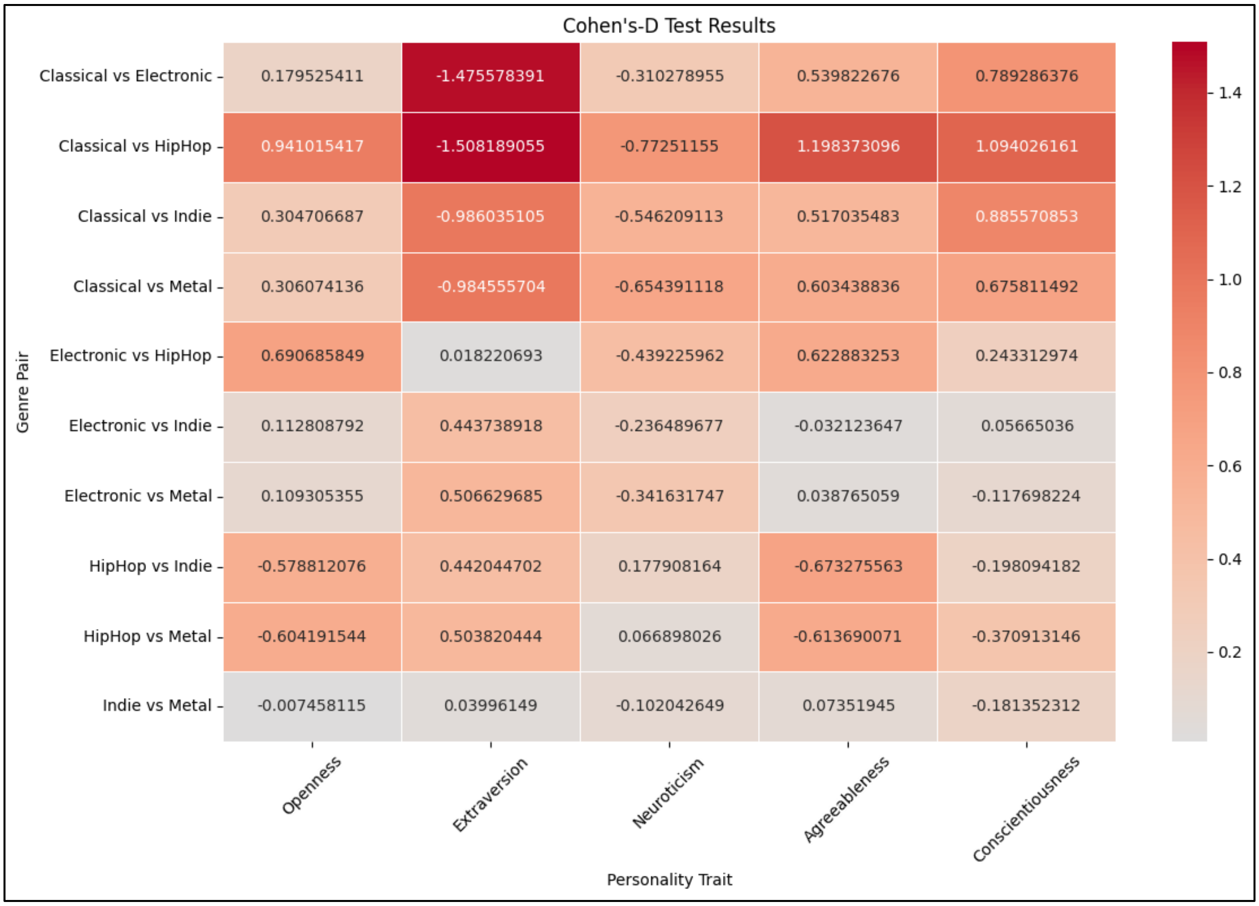

Figure 18: Cohen's-D test results heatmap

10.4. The Other Way Around: Predicting Genre by Personality

Although this was not the primary goal of our research, we undertook the challenging task of predicting a user's musical preference based solely on a five-dimensional personality vector derived from their music-unrelated text. Despite the difficulty inherent in mapping personality traits to specific musical tastes, we achieved encouraging results that met the random baseline, demonstrating some level of predictive power.

To complete the picture, we had one final experiment, predicting musical genre given a personality probability vector. To perform this test, we assigned a new "personality vector" for every user, which is a five-feature vector concluded by the user's probability to score high in each trait by order: OPN, EXT, NEU, AGR, CON. We then split the dataset to test (0.2) and train (0.8) sets and trained a logistic regression model which achieved an accuracy of 0.424, hitting the random baseline of 0.2.

The classification report in Table 14 presents the performance metrics for the logistic regression model trained to predict musical genres based on personality vectors. The model achieved an overall accuracy of 42%, slightly above the baseline of random guessing (20% for five classes). The highest performance was observed for the "classical music" genre, with a precision of 0.54, recall of 0.76, and F1-score of 0.63. However, the model struggled with "hiphopheads" and "Metal," reflected in notably low recall and F1 scores. These results indicate that while the model can somewhat distinguish between genres, there is substantial room for improvement in its predictive capability across different musical categories.

	precision	recall	f1-score	Test Size
metal	0.31	0.18	0.23	178
classicalmusic	0.54	0.76	0.63	221
electronicmusic	0.37	0.24	0.29	177
indieheads	0.38	0.76	0.51	201
hiphopheads	0.43	0.02	0.04	160
accuracy			0.42	937
macro AVG	0.41	0.39	0.34	937
weighted AVG	0.41	0.42	0.36	937

Table 14: Classification Report of logistic regression model predictions – from personality to genre

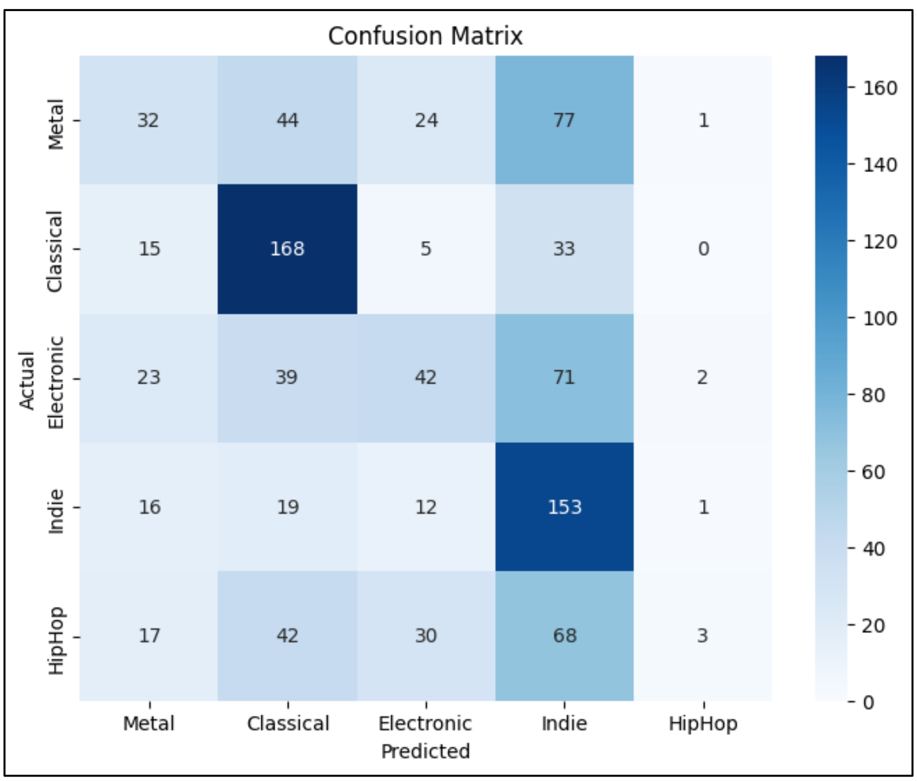

Figure 19: Confusion Matrix of logistic regression model predictions – from personality to genre

This model's predictions are also plotted in Figure 19, which highlights that it performs well in predicting Classical and Indie genres with most instances correctly classified. However, significant misclassifications occur, particularly with Metal and Hip-Hop which are often confused with Indie fans. These results suggest that while the model has some predictive power, especially for Classical and Indie, it exhibits considerable difficulty in distinguishing between certain genres, indicating a need for further refinement in feature representation or model complexity.

11. Conclusion and Future Work

This paper presents a comprehensive analysis of the relationship between musical preferences and personality traits as those are extracted from textual data from social media. By leveraging advanced natural language processing (NLP) techniques and large language models (LLMs), we revealed significant associations between personality traits, as defined by the Big Five model, and five distinct musical genres. Notably, Classical music enthusiasts exhibited higher agreeableness and lower extraversion, while Hip-Hop fans displayed lower agreeableness and higher neuroticism. These findings contribute valuable insights to the growing body of research at the intersection of psychology and computational linguistics, shedding light on how personality characteristics manifest in musical tastes.

However, this research is not without its limitations. The reliance on Reddit as a data source introduces potential biases, as the platform's demographic is skewed toward younger males, which could influence the generalizability of the findings. Additionally, the assumption that the most active subscribers of genre-specific subreddits are representative of fans of those genres is a strong (albeit not unreasonable) assumption, as not all subscribers may identify strongly as the genre's fans. Furthermore, confounding variables such as age and gender were not directly accounted for, which could have impacted the results. However, the dataset's large scale should provide a higher level of generality and overcome the potential demographic, age, or gender biases. Furthermore, past works have already taken high activity in a topic-related subreddit as a user property, identifying users with the subreddit's properties [84].

Looking forward, several avenues for future research emerge. One promising direction is the development of a more sophisticated classifier, potentially based on BERT or similar transformer models, fine-tuned on the dataset to enhance the accuracy of personality trait predictions. Expanding the study across multiple social media platforms could provide insights into whether the observed relationships hold across different user demographics. Additionally, integrating other personality models, such as the HEXACO, may uncover further correlations with musical preferences. There is also potential to leverage the findings in developing personalized music recommendation systems that factor in users' predicted personality traits, thereby improving the relevance of recommendations. Furthermore, refining the dataset generated using LLMs by incorporating manually assigned labels could further enhance prediction accuracy, offering a more reliable foundation for future research.

The dataset curated for this research represents a novel contribution to the field, providing a rich resource for further exploration of the interplay between language, personality, and musical preferences. The LLM-generated dataset of texts labeled with high and low presence of personality traits allows for the training of advanced models on an unbiased and unique textual dataset. This also strengthens previous work demonstrating that LLMs can reveal personality traits in their generated texts, opening new paths for research based on LLM-generated datasets. Specifically, texts generated by LLMs that exhibit different levels of personality traits offer promising avenues for further exploration. The five logistic regression models developed in this study offer a foundation for future work, but there is potential to expand upon this by integrating more advanced machine-learning techniques.

Ultimately, this research opens up new possibilities for understanding how our personalities shape, and are shaped by, the music we engage with, offering a compelling intersection of computational methods and psychological inquiry.

12. References

[1] Chávez, C. (n.d.). Our language is the reflection of ourselves. A language is an exact reflection of the character and growth of its speakers. Retrieved from https://ufw.org/research/history/education-heart-cesar-chavez-words/

[2] Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of cross-cultural psychology*, *38*(2), 173-212. https://doi.org/10.1177/0022022106297299

[3] Stajkovic, A. D., Bandura, A., Locke, E. A., Lee, D., & Sergent, K. (2018). Test of three conceptual models of influence of the big five personality traits and self-efficacy on academic performance: A meta-analytic path-analysis. Personality and individual differences, 120, 238-245. https://doi.org/10.1016/j.paid.2017.08.014

[4] Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. Journal of personality, 90(2), 222-255. https://doi.org/10.1111/jopy.12663

[5] Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychological bulletin*, *136*(5), 768. https://psycnet.apa.org/doi/10.1037/a0020327

[6] Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The big five personality factors and personal values. *Personality and social psychology bulletin*, *28*(6), 789-801. https://doi.org/10.1177/0146167202289008

[7] Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences*, *116*, 265-273. https://doi.org/10.1016/j.paid.2017.04.061

[8] Flannery, M. B., & Woolhouse, M. H. (2021). Musical preference: Role of personality and music-related acoustic features. *Music & Science*, *4*, 20592043211014014. https://doi.org/10.1177/20592043211014014

[9] Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality: Evidence from active listening and Facebook likes. Psychological science, 29(7), 1145-1158. https://doi.org/10.1177/0956797618761659

[10] Ruth, N., & Müllensiefen, D. (2020). Associations between musical preferences and personality in female secondary school students. *Psychomusicology: Music, Mind, and Brain*, *30*(4), 202. https://psycnet.apa.org/doi/10.1037/pmu0000267

[11] Kopacz, M. (2005). Personality and music preferences: The influence of personality traits on preferences regarding musical elements. *Journal of music therapy*, *42*(3), 216-239. https://doi.org/10.1093/jmt/42.3.216

[12] Herrera, L., Soares-Quadros Jr, J. F., & Lorenzo, O. (2018). Music preferences and personality in Brazilians. *Frontiers in Psychology*, *9*, 1488. https://doi.org/10.3389/fpsyg.2018.01488

[13] Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2020). Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, *53*(4), 2313-2339. https://link.springer.com/article/10.1007/s10462-019-09770-z

[14] Hu, R., & Pu, P. (2009, October). Acceptance issues of personality-based recommender systems. In *Proceedings of the third ACM conference on Recommender systems* (pp. 221-224). https://doi.org/10.1145/1639714.1639753

[15] Christodoulou, E., Gregoriades, A., Herodotou, H., & Pampaka, M. (2022, September). Extracting User Preferences and Personality from Text for Restaurant Recommendation. In *ORSUM@ RecSys*. https://ceur-ws.org/Vol-3303/paper9.pdf

[16] Ferwerda, B., Tkalcic, M., & Schedl, M. (2017). Personality traits and music genre preferences: how music taste varies over age groups. In *1st Workshop on Temporal Reasoning in Recommender Systems (RecTemp) at the 11th ACM Conference on Recommender Systems, Como, August 31, 2017.* (Vol. 1922, pp. 16-20). CEUR-WS. https://www.diva-portal.org/smash/get/diva2:1141105/FULLTEXT02

[17] Greenberg, D. M., & Rentfrow, P. J. (2017). Music and big data: a new frontier. *Current opinion in behavioral sciences*, *18*, 50-56. https://doi.org/10.1016/j.cobeha.2017.07.007
[18] Tupes, E. C., & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of personality*, *60*(2), 225-251. https://doi.org/10.1111/j.1467-6494.1992.tb00973.x
[19] Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), Review of Personality and Social Psychology (Vol. 2, pp. 141-165). Beverly Hills, CA: Sage. https://projects.ori.org/lrg/PDFs_papers/universals.lexicon.81.pdf
[20] John, O. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality/Guilford. https://psycnet.apa.org/record/1999-04371-004
[21] Galton, F. (1884). Measurement of character. fortnightly review, 36, 179-185. *Galton17936Fortnightly Review1884*. https://galton.org/bib/JournalItem.aspx_action=view_id=133
[22] Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. Journal of personality and social psychology, 77(6), 1296–1312. https://doi.org/10.1037//0022-3514.77.6.1296
[23] Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & emotion*, *10*(6), 601-626. https://doi.org/10.1080/026999396380079
[24] Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, *30*, 457-500. https://dl.acm.org/doi/10.5555/1622637.1622649
[25] Greenberg, D., Baron-Cohen, S., Stillwell, D., Kosinski, M. & Rentfrow, P. (2015). Musical Preferences are Linked to Cognitive Styles. https://doi.org/10.1371/journal.pone.0131151
[26] Kosan, M. A., Karacan, H., & Urgen, B. A. (2022). Predicting personality traits with semantic structures and LSTM-based neural networks. *Alexandria Engineering Journal*, *61*(10), 8007-8025. https://doi.org/10.1016/j.aej.2022.01.050
[27] Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011, October). Our twitter profiles, our selves: Predicting personality with twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing* (pp. 180-185). IEEE. http://dx.doi.org/10.1109/PASSAT/SocialCom.2011.26
[28] Wang, Y. (2015). Understanding personality through social media. *Journal of Psychology*. https://nlp.stanford.edu/courses/cs224n/2015/reports/6.pdf
[29] Bawa, G., Pakira, K., & Sharma, S. (2018). Extraction and Use of Personality Traits from Written Commentary. 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 1332-1335. https://doi.org/10.1109/CSCI46756.2018.00258
[30] Majumder, Navonil & Poria, Soujanya & Gelbukh, Alexander & Cambria, Erik. (2017). Deep Learning-Based Document Modeling for Personality Detection from Text. IEEE Intelligent Systems. 32. 74-79. 10.1109/MIS.2017.23. http://dx.doi.org/10.1109/MIS.2017.23
[31] Sun, Xianguo & Liu, Bo & Cao, Jiuxin & Luo, Junzhou & Shen, Xiaojun. (2018). Who Am I? Personality Detection Based on Deep Learning for Texts. 1-6. 10.1109/ICC.2018.8422105. http://dx.doi.org/10.1109/ICC.2018.8422105
[32] Heci Zheng and Chunhua Wu. Predicting personality using Facebook status based on semi-supervised learning. In ACM International Conference Proceeding Series, volume Part F148150, pages 59–64, New York, New York, USA, 2019. Association for Computing Machinery. ISBN 9781450366007. doi: 10.1145/3318299.3318363. URL http://dl.acm.org/citation.cfm?doid=3318299.3318363.
[33] Ren, Z., Shen, Q., Diao, X., & Xu, H. (2021). A sentiment-aware deep learning approach for personality detection from text. Information Processing & Management, 58(3), 102532. http://dx.doi.org/10.1016/j.ipm.2021.102532
[34] Sirasapalli, J. J., & Malla, R. M. (2023). A deep learning approach to text-based personality prediction using multiple data sources mapping. Neural Computing and Applications, 35(28), 20619-20630. https://link.springer.com/article/10.1007/s00521-023-08846-w

[35] Yang, K., Lau, R. Y., & Abbasi, A. (2023). Getting personal: A deep learning artifact for text-based measurement of personality. Information Systems Research, 34(1), 194-222. https://doi.org/10.1287/isre.2022.1111
[36] Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, *108*(6), 934. https://psycnet.apa.org/doi/10.1037/pspp0000020
[37] Peters, H., & Matz, S. (2024). Large language models can infer psychological dispositions of social media users. *PNAS Nexus*, pgae231. https://arxiv.org/pdf/2309.08631
[38] Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Romero, P., ... & Matarić, M. (2023). Personality traits in large language models. *arXiv preprint arXiv:2307.00184*. https://doi.org/10.48550/arXiv.2307.00184
[39] Cattell, R. B. & Anderson, J. C. (1953). The measurement of personality and behavior disorders by the I.P.A.T. Music Preference Test. Journal of Applied Psychology, 37, 446- 454. http://176.9.41.242/doc/psychology/1953-cattell.pdf
[40] Cattell, R. B. & Saunders, D. R. (1954). Musical preference and personality diagnosis: I. A. factorization of one hundred and twenty cases. Journal of Social Psychology, 39(1), 3-24. https://gwern.net/doc/psychology/personality/1954-cattell.pdf
[41] Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, *84*(6), 1236. https://psycnet.apa.org/doi/10.1037/0022-3514.84.6.1236
[42] Bonneville-Roussy, Arielle; Rentfrow, Peter J.; Xu, Man K.; Potter, Jeff (2013). "Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood". Journal of Personality and Social Psychology. 105 (4): 703–717. doi:10.1037/a0033770. PMID 23895269.
[43] Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: a five-factor model. *Journal of personality and social psychology*, *100*(6), 1139. https://psycnet.apa.org/doi/10.1037/a0022406
[44] Greenberg, D. M., Wride, S. J., Snowden, D. A., Spathis, D., Potter, J., & Rentfrow, P. J. (2022). Universals and variations in musical preferences: A study of preferential reactions to Western music in 53 countries. *Journal of personality and social psychology*, *122*(2), 286. https://psycnet.apa.org/doi/10.1037/pspp0000397
[45] Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). "Just the way you are": Linking music listening on Spotify and personality. *Social Psychological and Personality Science*, *12*(4), 561-572. https://doi.org/10.1177/1948550620923228
[46] Sust, L., Stachl, C., Kudchadker, G., Bühner, M., & Schoedel, R. (2023). Personality computing with naturalistic music listening behavior: Comparing audio and lyrics preferences. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.75214
[47] Veltri, T. M. (2016). Personality correlates of preferences for music structure and emotion. Unpublished master's thesis, The University of Sheffield. http://dx.doi.org/10.13140/RG.2.1.2340.8721
[48] Neuman, Y., Perlovsky, L., Cohen, Y., & Livshits, D. (2016). The personality of music genres. *Psychology of Music*, *44*(5), 1044-1057. http://dx.doi.org/10.1177/0305735615608526
[49] https://arc.net/l/quote/byjckxas
[50] https://arc.net/l/quote/eyvgguub
[51] https://arc.net/l/quote/orppongd
[52] https://arc.net/l/quote/stnkyukc
[53] https://arc.net/l/quote/qhyjnpuz
[54] https://arc.net/l/quote/uyxzlngm
[55] https://arc.net/l/quote/xdijmcyi
[56] "SentenceTransformers" Python library documentation - sbert.net
[57] Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., ... & Wei, F. (2022). Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*. https://arxiv.org/pdf/2212.03533
[58] https://www.verywellmind.com/the-big-five-personality-dimensions-2795422

[59] The Big 5 Personality Traits. (2023). https://psychcentral.com/lib/the-big-five-personality-traits Medically reviewed by Marney A. White, PhD, MS, Psychology — By John M. Grohol, Psy.D. and Janelle Cox
[60] Big Five Personality Traits: The 5-Factor Model Of Personality. (2023). https://www.simplypsychology.org/big-five-personality.html
[61] Costa, P. T., & McCrae, R. R. (1999). A five-factor theory of personality. *Handbook of personality: Theory and research*, *2*(01), 1999. https://www.researchgate.net/publication/284978581_A_five-factor_theory_of_personality
[62] The Personality of Academic Majors | Scientific American. (2023). https://www.scientificamerican.com/article/the-personality-of-academic-majors/
[63] Romeo, M., Hernández García, D., Han, T., Cangelosi, A., & Jokinen, K. (2021). Predicting apparent personality from body language: benchmarking deep learning architectures for adaptive social human–robot interaction. *Advanced Robotics*, *35*(19), 1167-1179. https://doi.org/10.1080/01691864.2021.1974941
[64] Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye movements during everyday behavior predict personality traits. *Frontiers in human neuroscience*, *12*, 328195. https://doi.org/10.3389/fnhum.2018.00105
[65] https://ipip.ori.org/
[66] Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013). Workshop on computational personality recognition: Shared task. In Proceedings of the International AAAI Conference on Web and social media (Vol. 7, No. 2, pp. 2-5).
[67] https://sites.google.com/michalkosinski.com/mypersonality
[68] https://github.com/fuguixing/psychology-insights-frontend
[69] https://github.com/fuguixing/psychology-insights-synapse-training
[70] https://farm2.user.srcf.net/research/personality/recognizer.html
[71] François Mairesse, Marilyn Walker, Matthias Mehl and Roger Moore. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. Journal of Artificial Intelligence Research (JAIR), 30, pages 457-500, 2007.
[72] Cambridge University, The Psychometrics Centre. (2023). Apply Magic Sauce [Computer software]. Retrieved from https://applymagicsauce.com
[73] https://github.com/jkwieser/personality-prediction-from-text/tree/master
[74] Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. IEEE intelligent systems, 32(2), 74-79.
[75] jbei/scikit-learn docker image, based on Python 3 image with NumPy, SciPy, scikit-learn, and python-libsbml pre-installed: https://hub.docker.com/r/jbei/scikit-learn
[76] Open-AI's Chat-GPT: https://chatgpt.com/
[77] Google's Bard, Gemini today: https://gemini.google.com/app
[78] Bing Co-Pilot: https://www.bing.com/chat
[79] Zulqarnain, M., Ghazali, R., Ghouse, M. G., & Mushtaq, M. F. (2019). Efficient processing of GRU based on word embedding for text classification. JOIV: International Journal on Informatics Visualization, 3(4), 377-383. http://dx.doi.org/10.30630/joiv.3.4.289
[80] Jung, S., Lee, J., & Kim, J. (2018, October). Learning to embed semantic correspondence for natural language understanding. In Proceedings of the 22nd Conference on Computational Natural Language Learning (pp. 131-140). https://doi.org/10.18653/v1/K18-1013
[81] Meta's Dragon+ BEIR benchmark: https://github.com/facebookresearch/dpr-scale/tree/main/dragon
[82] intfloat/e5-large-v2 BEIR Benchmark result: https://github.com/microsoft/unilm/tree/master/e5#english-pre-trained-models
[83] MTEB benchmark leaderboard: https://huggingface.co/spaces/mteb/leaderboard
[84] Rabinovich, E., Tsvetkov, Y., & Wintner, S. (2018). Native language cognate effects on second language lexical choice. Transactions of the Association for Computational Linguistics, 6, 329-342.

13. Appendix

This work is based on resources available on the social network Reddit.

1. 21 subreddits used at the initial validation phase:
   ['r/classicalmusic', "r/opera", "r/ambientmusic", "r/chillout", "r/deephouse", "r/DnB", "r/dubstep", "r/EDM", "r/electronicdancemusic", "r/rock", "r/indieheads", r/electronicmusic", "r/blues", "r/folk", "r/Metal", "r/punk", "r/hiphopheads", "r/hiphop101", "r/rap", "r/Jazz", "r/reggae"]
2. 5 subreddits selected for research:
   ['r/indieheads', 'r/electronicmusic', 'r/hiphopheads', 'r/Metal', 'r/classicalmusic']
3. 142 subreddits that were excluded to protect the detailed data's validity:
   ["trap", "OutlawCountry", "Ska", "SoundsVintage", "indie_rock", "MelodicDeathMetal", "PsychedelicRock", "Deathcore", "GarageRock", "KoreanRock", "Noisey", "PopHeads", "Punk_Rock", "Synthpop", "ClassicRock", "Grime", "MetalMemes", "MusicTheory", "EmoRevival", "PunkRockers", "AlternativeRock", "ElectronicJazz", "Bluegrass", "ChristianMusic", "PsychRock", "DeathMetal", "Rockabilly", "Kpopthoughts", "HipHopCollabs", "RocknRoll", "ElectroBlues", "Djent", "Trance", "SwingMusic", "hiphopheads HOF", "FolkPunk", "PopRock", "PopPunk", "EmoScreamo", "MelodicMetalcore", "Rockband", "Chiptunes", "Albums", "NewWave", "IndieElectronica", "Metallica", "Psybient", "ProtestTheHero", "EmoForever", "PopPunkEmo", "hiphopheads", "EDM", "rock", "metal", "popheads", "country", "punk", "indieheads", "rnbheads", "electronicmusic", "ClassicalMusic", "hiphop101", "alternativerock", "techno", "jazz", "kpop", "Blues", "Musicthemetime", "Soulies", "reggae", "MusicAlbums", "Emo", "GratefulDead", "PostHardcore", "PowerMetal", "hiphoptruth", "futurebeats", "progrockmusic", "Vaporwave", "musicals", "synthwave", "hiphop101", "Metalcore", "PopPunkers", "Dubstep", "IndieFolk", "psytrance", "hiphopimages", "lofihiphop", "OutRun", "industrialmusic", "IndieWok", "Coffeeshopvibes", "Electrohouse", "stonerrock", "swinghouse", "hiphophub", "SymphonicMetal", "Emo_irl", "progmetal", "electronicmusic", "Jazz", "classical_circlejerk", "Baroque", "Mozart", "Chopin", "DnB", "Techno", "skrillex", "dubstep", "deephouse", "KendrickLamar", "Kanye", "Hiphopcirclejerk", "Eminem", "MacMiller", "HipHopGoneWild", "HipHopShit", "indieheadscirclejerk", "radiohead", "arcticmonkeys", "Kanye", "beatles", "TaylorSwift", "TheKillers", "RedHotChiliPeppers", "gorillaz", "Jazz", "lorde", "blur", "TheCure", "radioheadcirclejerk", "gratefuldead", "beatlescirclejerk", "Muse", "Blink182DavidBowieMegadeth", "BlackMetalMemes", "pinkfloyd", "Coldplayrem", "DaftPunk", "BlackMetal", "MetalPorn"]
4. Community_summary.xlsx: summary of counts of all the posts and comments in all Reddit, done by the musical genres' fans, https://github.com/eliranshemtov/Musical-Preferences-And-Textual-Expression/blob/main/analysis/community_summary.xlsx
5. Redditors_activity_summary.csv: user-focused table, listing the top 10 subreddits per Redditor, https://github.com/eliranshemtov/Musical-Preferences-And-Textual-Expression/blob/main/analysis/redditors_activity_summary.csv
6. Community_summary_filtered.xlsx: community_summary.xlsx after filtering of empty posts and activity from banned subreddits, https://github.com/eliranshemtov/Musical-Preferences-And-Textual-Expression/blob/main/analysis/community_summary_filtered.xlsx
7. Redditors_activity_summary_filtered.csv: redditors_activity_summary.csv after filtering of empty posts and activity from excluded subreddits, https://github.com/eliranshemtov/Musical-Preferences-And-Textual-Expression/blob/main/analysis/redditors_activity_summary_filtered.csv
8. Personality Recognizer - a Java command-line application that reads a set of text files and computes estimates of personality scores along the Big Five dimensions. http://farm2.user.srcf.net/research/personality/recognizer.html
9. Focus on the 5 selected musical genres' communities Jupyter notebook in the research's git repository: https://github.com/eliranshemtov/Musical-Preferences-And-Textual-Expression/blob/main/Step2-Focus-5-Genres.ipynb
10. Focus on the top 1000 authors in each community Jupyter notebook in the research's git repository: https://github.com/eliranshemtov/Musical-Preferences-And-Textual-Expression/blob/main/Step3-Focus-Top-1000-Authors.ipynb
11. The Big Five personality traits definitions and raw LLM generated datasets in the research's git repository: https://github.com/eliranshemtov/Musical-Preferences-And-Textual-Expression/blob/main/analysis/llm-dataset-generation/traits-definitions.xlsx

12. LLM Main Prompts used for dataset generated by LLM models: https://mailmtaac-my.sharepoint.com/:f:/g/personal/eliranse_mta_ac_il/En3yqmzyZTlPnrkcdzrDOUQBuTR9unVZMfEqTssGrryE9w?e=GAjUgr
13. Appendix Table 1:

	myPersonality (250 users, 9,917 updates)			**Essays (2,468 users, 2,468 documents)**		
Prediction type	Discrete and continuous, separately			Discrete only		
	Mean	STD Deviation	Min-max	Mean	STD Deviation	Min-max
No. of posts per user	39.7	43.6	1-223	1.0	0	1-1
EXT	3.29	0.86	1.33-5.00	0.52	0.50	0-1
NEU	2.63	0.78	1.25-4.75	0.50	0.50	0-1
AGR	3.60	0.67	1.65-5.00	0.53	0.50	0-1
CON	3.52	0.74	1.45-5.00	0.51	0.50	0-1
OPN	4.07	0.58	2.25-5.00	0.52	0.50	0-1
No. of words per post	14.74	12.76	1-113	663.1	267.5	34-3,836
No. of nouns per post	2.81	2.68	0-37	80.66	34.5	5-294
No. of verbs per post	0.81	1.24	0-12	41.10	19.1	1-178
No. of adjectives per post	1.02	1.35	0-36	37.78	16.8	2-165
No. of adverbs per post	0.99	1.39	0-15	63.66	29.2	3-290
No. of concepts per post	11.62	6.80	0-39	45.74	3.4	22-51

Appendix Table 1: The "Gold-Standard" datasets properties [Essays and myPersonality]

14. Appendix Table 2:

Trait		Training Set, Generated by Chat-GPT		Test Set, Generated by Google's Bard	
		High	Low	High	Low
OPN	Sample	As the sun dipped below the horizon, I found myself lost in the vivid hues of the evening sky. My mind wandered, contemplating the mysteries of the cosmos and the infinite possibilities that lay beyond our earthly realm. In that moment, I realized the beauty of embracing the unknown, a sentiment that fuels my insatiable curiosity and thirst for discovery. Whether it's exploring uncharted territories or delving into the intricacies of abstract art, my open mind is a canvas eager to be painted with the brushstrokes of imagination and innovation.	Predictability is my preference. The unknown and the unfamiliar create a sense of discomfort that I actively avoid. Imagination and creativity seem like elusive concepts, and I find fulfillment in the stability of routine. Change may excite some, but for me, it's an unwelcome disruption.	The museum at 3 AM isn't just quiet - it hums with forgotten stories. Moonlight slants through stained glass, painting the Egyptian sarcophagus with an eerie glow. Every shadow seems to shift, hieroglyphs in constant flux. I'm not afraid, just exhilarated by the whispers of history, these lives tangled in dust motes and moonbeams. Maybe tonight, under the watchful gaze of ancient gods, I'll crack their code, decipher the echoes of their laughter and pain.	My morning routine is like clockwork. Coffee at 7:00, newspaper by 7:30, out the door for work by 8:00. I wouldn't have it any other way. New things just mess me up. Remember that time I tried a new Thai restaurant? My stomach was in knots for days. I'll stick to my meatloaf, thanks very much.
	Total	149	215	95	93
	Unique	143	129	51	93
	Mean WC	65.52	60.3	60.43	64.59
	Min WC	40	39	41	40
	Max WC	105	100	98	102
	Sample	Hey there! I absolutely thrive in social settings – the more	Social gatherings always leave me drained. It's not	Organized a clothing swap party with my friends! It was a night of	In a world that values loud voices and constant

Trait		Training Set, Generated by Chat-GPT		Test Set, Generated by Google's Bard	
		High	Low	High	Low
EXT		people, the merrier! It's like every conversation is an opportunity to learn something new or share a laugh. I just love being in the middle of it all, soaking up the energy and excitement around me. Whether it's a party, a networking event, or just a casual get-together, count me in!	that I dislike people, but the constant buzz of conversation exhausts me. I find solace in quiet moments alone, where I can recharge my energy and gather my thoughts. Small talk feels superficial to me, and I struggle to engage in it genuinely. Instead, I prefer deeper conversations that delve into meaningful topics. Being the center of attention is my worst nightmare; I much prefer blending into the background and observing rather than being in the spotlight.	fashion, fun, and sustainability. We traded pre-loved clothes, accessories, and even books. It was a fantastic way to refresh our wardrobes without contributing to fast fashion waste. Plus, it was a blast catching up with friends and discovering hidden fashion gems! #ClosetCleanoutChallenge	self-promotion, I find strength in the quiet act of listening. When someone chooses to confide in me, I become a vessel for their words, absorbing their emotions without judgment. This space for open and honest communication fosters a deeper connection, even without speaking a single word.
	Total	176	359	109	109
	Unique	167	286	105	106
	Mean WC	80.92	69.3	86.21	107.77
	Min WC	38	34	55	55
	Max WC	199	142	191	253
NEU	Sample	Every day is a struggle, a relentless marathon through the murky depths of my own mind. It's like I'm constantly wading through a swamp of negativity, each step dragging me further down into the mire of despair. I try to shake off the oppressive weight of my thoughts, but they cling to me like leeches, draining me of energy and hope.	Exploring the wonders of nature has always been my sanctuary. There's something deeply calming about immersing oneself in the tranquility of a forest or by the gentle lapping of waves at the shore. Each time I venture out into the wilderness, I'm reminded of how fortunate I am to experience such serenity and beauty. It's as if all worries melt away, leaving behind a profound sense of peace and contentment.	The silence is suffocating. My fingers hover over the keyboard, a blank page mocking me. Every idea I have feels derivative, uninspired. Did I ever have talent, or was it just a cruel trick of the light? My stomach clenches, the anxiety a familiar unwelcome guest. Maybe this deadline is too tight, everyone knows I work best under pressure, right? But what if this time the pressure crushes me? What if this is the project that exposes me for the fraud I truly am?	I try to be mindful of my environmental impact. It's not about feeling guilty or overwhelmed – it's about making small, conscious choices. I recently switched to a green energy provider for my apartment, and it felt great knowing I was making a positive difference, however small. Plus, who doesn't love a lower electricity bill?
	Total	192	187	187	148
	Unique	173	181	177	148
	Mean WC	65.76	60.7	69.09	75.29
	Min WC	38	31	20	42
	Max WC	110	110	133	126
AGR	Sample	Protecting our planet and its delicate ecosystems has always been a priority close to my heart. Every action we take, big or small, contributes to the health of our environment. Whether it's reducing our carbon footprint or advocating for renewable energy sources, we have the power to make a positive impact.	It's utterly baffling how some people can't seem to handle the simplest tasks without constant hand-holding. I mean, come on, do I look like your personal assistant? Get it together and figure it out yourself for once. I've got my own stuff to deal with, and I certainly don't have time to babysit grown adults who can't take responsibility for themselves.	Historical figures like Jane Austen, with her witty social commentary and timeless stories of love and marriage, continue to be relevant today. Her ability to capture the nuances of human relationships and societal expectations, often with a touch of humor, offers insightful perspectives that resonate across generations.	This whole fascination with self-improvement is a waste of time. People are who they are. You can't polish a turd, as they say. Those self-help gurus are just selling snake oil to people who are too weak to accept themselves. Focus on your strengths and exploit the weaknesses of others. That's the real path to success.
	Total	179	170	149	149

Trait		Training Set, Generated by Chat-GPT		Test Set, Generated by Google's Bard	
		High	Low	High	Low
	Unique	157	164	147	149
	Mean WC	87.19	62.25	83.54	73.25
	Min WC	43	38	47	43
	Max WC	171	141	122	118
CON	Sample	Embracing innovation is key to addressing the complex challenges of our time. I believe in the power of creativity and forward thinking to drive progress and transformation. By harnessing new technologies and ideas, we can develop innovative solutions that improve lives and shape a better future for all.	You ever try to stick to a schedule? It's like trying to herd cats, man. I'll set one up, then something shiny catches my eye, and poof, there goes the plan. But hey, spontaneity is the spice of life, right? Who needs boring old routines?	My desk is like a mini-universe, organized by function. Textbooks stacked by subject on the left, notebooks color-coded by class in the center, and a planner with a weekly schedule meticulously mapped out on the right. Every morning, it's a quick glance at the plan for the day, packing my bag based on what classes I have, and double-checking my homework before heading out the door. I thrive on routine, and a little structure keeps me focused on the ultimate goal – acing that final exam and getting into college.	They're talking about colonizing the moon next. Seriously? The moon? It's a giant rock with no air or water. Who wants to live in a giant, dusty cheese wheel? Sounds like a nightmare.
	Total	179	159	139	149
	Unique	157	149	139	149
	Mean WC	78.73	63.24	70.94	56.61
	Min WC	47	32	40	27
	Max WC	146	123	151	123

Appendix Table 2: The LLM generated train and test sets with high vs low presence of each personality trait, Samples, and statistics.