Examination of the Use and Misuse of Psychiatric Terminology in Public Discourse Using Large Language Models

MSc. Thesis

Viola Louisa Kahlert

University of Twente
Faculty of Behavioural Management Sciences (BMS)
Department of Technology, Human and Institutional Behaviour (HIB)
Section of Psychology and Health Technology

Supervision and Examination Committee
Dr. J. Piano Simoes
Dr. J.A. Koelen

July 18, 2024

Abstract

Social media made public discourse accessible but also increased online incivility. Mental health stigma is a frequent problem linked to the derogatory misuse of psychiatric terminologies such as neurotic and neuroticism. This study investigated the use of these terms on Twitter examining frequency, context and offensiveness. It was hypothesized that engagement would differ between the two terms, with neurotic being used in a more pejorative manner. Moreover, it was proposed that offensive use increased over time, showing different temporal trends in the discourse involving neurotic compared to neuroticism.
This observational text mining study analysed 427,027 English tweets mentioning the term ‘neurotic’ and 33,282 mentioning the term ‘neuroticism’ posted between January 2015 and August 2021. The study utilised the roBERTa for offensive language detection, BERTopic for topic modelling, and ChatGPT 3.5 for labelling topics, along with statistical analyses to compare interaction frequencies and offensive content over time. The results revealed that the term neurotic is more commonly used than neuroticism, with 31% of neurotic-mentioning tweets being offensive compared to 8% for neuroticism. Engagement was higher for tweets with the keyword neuroticism. Topic modelling identified 89 themes for neurotic and 71 for neuroticism, indicating a broader range of contexts, including more derogatory uses for neurotic. The findings highlight the need for educational interventions and stricter policies to promote appropriate use of psychiatric terminology. Furthermore, ongoing research is crucial to monitor and address misuse of these terms to reduce stigma and improve mental health literacy.

Keywords. public discourse, Twitter, text mining, psychiatric terminology, offensive language detection

Contents

Examination of the Use and Misuse of Psychiatric Terminology in Public Discourse Using Large Language Models

Social media has fundamentally transformed public discourse, rendering it more accessible, integrative, and democratic – a phenomenon comparable to the revolutionary impact of the printing press during the medieval times. In recent years, the expansion of social media platforms led to an active engagement of 61.4 % of the global population (Kemp, 2023). This widespread adoption of social media across geographical boundaries fostered global communication and collaboration (Hizbul Khootimah Azzaakiyyah, 2023). Today, every social media user can share their opinions and thoughts online with a global audience and respond and engage with the content created by others. Through this, social media facilitated cultural exchange, social movement, individual expression, and global awareness and promoted social support communities (Hizbul Khootimah Azzaakiyyah, 2023).

However, social interactions are suffering under a rising trend of online incivility (Antoci et al., 2019). 41% of U.S. citizens experienced aggression, harassment and hate speech online. Hate speech often aims at insulting or denigrating social groups (Schmid et al., 2022). Particularly, discussions surrounding sensitive or controversial topics such as politics and gender are prone to shift into personal attacks and provocative remarks. Social media platforms due to their lack of face-to-face interaction have a disinhibiting effect on the users since they cannot see the emotional reactions of whom they speak to (Antoci et al., 2019). Users' anonymity adds to the potential of cruelty in online communication. The anonymity and invisibility online results in users expressing their opinions more directly and impulsively, often refraining from revising their thoughts once they are posted (Suler, 2004). The societal norms, arguments, self-awareness, and restraint that play a key role in offline public discourse are absent online, leading to increased incivility (Antoci et al., 2019).

Stigmatising people based on their mental health problems is a common form of denigration (Arboleda-Flórez & Stuart, 2012). The stigma is based on discrimination, stereotypes, and prejudice (Corrigan & Watson, 2002). Particularly, psychotic disorders are highly stigmatized with affected people wrongly assumed to be aggressive and impulsive (Henderson et al., 2013). Mental illnesses are strikingly more stigmatised than physical illnesses (Corrigan & Watson, 2002). The stigma surrounding mental illness increases regardless of the improving mental health awareness (Hinshaw & Stier, 2008).

Individuals with mental illnesses are not only stigmatised, but the stigma adversely impacts their mental well-being (Benesch, 2014; Li et al., 2020). Seven out of ten individuals with a mental health condition are not in treatment as they are reluctant to seek out professional help after being exposed to stigma (Henderson et al., 2013). Moreover, stigma decreases self-esteem and social participation by eliciting fear and rejection (Kallivayalil & Enara, 2016; Naslund & Deng, 2021; Hinshaw & Stier, 2008).

Part of stigmatising language is the wrongful labelling of groups affected by mental health problems. A linguistic study by Hogeweg & Neuleman (2022) investigated the hurtfulness of labelling terms using examples of slurs, nouns, and adjective-noun combinations. The study revealed that there is no clear difference between neutral and non-neutral labels but that the conventional meaning and the context define the word's hurtfulness. Unlike nouns the offensiveness of an adjective can be amplified by adverbs (Padilla Cruz, 2019). For example, in the phrase 'She is painfully neurotic about cleanliness', the adverb painfully reinforces the derogatory notion of the word neurotic. This understanding of the impact of labels highlights how language targeting groups or their members has an enormous potential to promote intergroup conflicts (Carnaghi & Bianchi, 2017). Moreover, it explains why health-related stigma emerges particularly in social media with heightened incivility (Competiello et al., 2023). Although labels are associated with stigma, classifications are needed in mental health care. To avoid the stigma that results from categorical attributions spectrum and continuum models were introduced in mental health (Fernandez et al., 2022). Nonetheless, the careful use of psychiatric terminology remains relevant considering stigma in relation to semantics and syntax.

The term neurotic serves as an example in which mental health can be used as a stigmatising label. Neurotic labels individuals with high levels of neuroticism or symptoms of a neurosis highlighting its stigmatising nature (Arboleda-Flórez & Stuart, 2012; Britannica, 2019; Weed & Kwon, 2019). Two clinically reviewed articles confirm that the misuse of the term neurotic may be due to its resemblance with the word neurosis, an outdated diagnosis for otherwise uncategorisable behavioural anomalies (Chung, 2019; Gillette, 2022). The word neurosis was popularized with the advent of psychoanalysis, which ascribed sexual frustration as the underlying cause of neurotic symptoms (Badcock, 1992; Cassiello-Robbins et al., 2017).

Today, psychologists use the term neurotic in the context of neuroticism. Neuroticism describes a personality trait domain that indicates proneness to develop a broad variety of mental illnesses (Widiger & Oltmanns, 2017). It is considered one of the main personality dimensions within models such as the Big Five (Cassiello-Robbins et al., 2017; Widiger & Oltmanns, 2017). A person with elevated neuroticism is more likely to feel angry, depressed, anxious, threatened, irritable, or emotionally unstable and react sensitively to external stressors (Widiger & Oltmanns, 2017). Additionally, neuroticism is associated with adverse life events as well as several psychiatric and other medical comorbidities (Lahey, 2009). Individuals scoring high on neuroticism are prone to develop self-stigma which poses a vulnerability for suicidality (Klára Látalová et al., 2014; Szcześniak et al., 2021).

Despite its importance, the misuse of the word neurotic has not yet been explored in academic research. McGarry (1990) described the misuse in the British Journal of Psychiatry highlighting the existing awareness of this phenomenon. Nonetheless, a list of 50 selected psychiatric words was compiled that are regularly misused, including terms like denial, splitting or closure (Lilienfeld et al., 2015). However, the word neurotic was not discussed on the list. Possibly misusing the term is more common in the general public whereas the list was to clarify terminologies that are not correctly used in the psychiatric sector.

Social media's transformative role in public discourse, the rise of online incivility and the interrelated mental health stigma are well known. However, research lacks an understanding of how the public uses psychological concepts of neurotic and neuroticism. While the historical development in the psychiatric use of neurotic and neuroticism from the classification of psychopathology to a dimension of personality is documented, it is unclear how this shift is reflected in the discourse on social media. Filling this gap is crucial as the public use of such terms influences perceptions of mental health and potentially contributes to stigma. Addressing the gap, this study leverages advances in large language models to analyse a large corpus of text from naturalistic settings where interactions involving the terms can be evaluated. Specifically, the present study seeks to examine the use of the terms neurotic and neuroticism in public discourse with the following research questions and hypothesis:

RQ1: Which of the psychiatric terminologies, neurotic or neuroticism, prevails in public discourse?

H1: There is a difference between the amount of interaction with tweets mentioning neurotic compared to neuroticism.

RQ2: Is there a noticeable difference in public use of the terms neurotic and neuroticism?

H2: The label neurotic is more frequently utilized to denigrate in comparison to the term neuroticism.

RQ3: How has the use of the terms by the public developed over time?

H3a: The offensive use of both terms neurotic and neuroticism has increased over time.

H3b: Discourse involving the word neurotic has different temporal trends than surrounding neuroticism.

RQ4: What are the key themes discussed in the context of the terms neurotic and neuroticism?

Method

The methods to address the above-stated research questions and hypotheses were described according to the STROBE checklist to enhance their transparency and clarity (Elm et al., 2007). However, there is no specific checklist for text mining reports therefore the STROBE statement for cross-sectional studies was used as a template and adjusted when necessary. Ethical approval was obtained on the 13th of March 2024 from the BMS Ethics Committee of the University of Twente.

Study Design

The study was set up as an observational approach using text mining on two Twitter datasets. The naturalistic sample consisted of English tweets with the keywords: neurotic and neuroticism. The main steps consisted of scraping the tweets, data preprocessing, descriptive statistics, offensive language detection and topic modelling.

Setting and Study Size

First, the data was retrieved from Twitter using snscrape (JustAnotherArchivist, 2020) on the 25th of March 2023. The library snscrape requiring Python 3.8 or newer is a tool to download large sets of data from social media platforms. Two pipelines compiled datasets with tweets containing either the keyword neurotic or neuroticism. The datasets were saved into CSV files including tweets from the 13th of January 2015 up to the 31st of August 2021. The time frame ensured that the tweets covered polarising events such as Brexit, the COVID-19 pandemic and the 2016 Donald Trump election caused a great amount of controversial discourse on social media (Bisiada, 2022; Gorodnichenko et al., 2021). Additionally, the files entailed the counts of likes, retweets, replies, and quotes as well as the dates of the account creation and tweet posting, tweet IDs, usernames and profile descriptions. Only English tweets were scraped to make the subsequent analysis more feasible but sufficiently extensive as it is the most used language on Twitter (Alshaabi et al., 2021). The scraping process resulted in a collection of 427027 tweets with the term neurotic and 33282 with neuroticism.

Data Sources and Measurement

Twitter as a social media platform with a notable prevalence of incivility and low moderation compared to other platforms is particularly useful for examining the potentially insulting use of “neurotic” and “neuroticism” (Bączkowska, 2021). Next to Facebook, Twitter is considered a monopoly for public and unstructured discourse (Antoci et al., 2019; Bączkowska, 2021). From the first tweet in March 2006 until the year 2022 up to three trillion tweets have been posted (Pfeffer et al., 2023). The content of tweets tends to be rather straight to the point and displays emotions uninhibitedly which generates uncivil interactions.

Additionally, Twitter offers the option to express criticism via handles without direct confrontation (Bączkowska, 2021). Twitter is known as a platform where people share their unfiltered thoughts (Damanik, 2023). Due to its unfiltered representation of social interactions and the large quantity of data Twitter has been considered a sociocope that has been utilized in many studies to investigate numerous facets of communication (Pfeffer et al., 2023).

Statistical Methods and Variables

Twitter allows for analyses of naturalistic interactions of thousands of people every day, it contains big data calling for tools to analyse such an amount of data. A prominent method to analyse large amounts of unstructured text is text mining (Hassani et al., 2020). Text mining is a method to detect patterns in either unstructured or semi-structured textual data (Cai & Sun, 2009). Automated text analysis was impossible until the invention of Natural Language Processing (Chowdhary, 2020).

Text mining enables the analysis of large datasets and thus overcomes limitations such as the lack of representativeness as the computational power exceeds the human capacities to process text data (Heinke, 2023). Moreover, examining extensive datasets can unveil a greater number of themes and cover larger periods to uncover temporal developments. Due to the recent introduction of transformer-based text-mining software, typical linguistic features in Twitter such as irony, sarcasm, abbreviations and emojis can be comprehended by the computer, providing a more nuanced interpretation of the tweet content over traditional text-mining tools (Heinke, 2023).

BERT, short for Bidirectional Encoder Representations from Transformers, is an advanced Natural Language Processing (NLP) model that can read the entire sequence of words at once instead of reading it sequentially like traditional models (Devlin et al., 2019). BERT processes text bidirectionally, meaning it understands the context of a word based on the words surrounding it. The model was pre-trained using two tasks: the masked language model, where some words of a sentence were hidden and the model predicted them and the next sentence prediction, where the model learned the relationship between sentences. These pre-training tasks allow BERT to capture complex language patterns and relationships (Devlin et al., 2019). BERT and its optimizations such as BERTopic and roBERTa were foundational to the analysis of the large dataset, with details about implementation discussed below.

Descriptives (RQ1/H1)

To answer which psychiatric term entailing tweets is more frequent the number of tweets retrieved from each dataset can be compared but to get insights into which tweets generate more interactions descriptive statistics were calculated. This and subsequent analyses were conducted in Simple Linux Utility for Resource Management (Slurm) jobs that ran Python scripts (see Appendix A) on the High-Performance Computing (AM HPC) Cluster of the University of Twente (Van Corbach, 2020). Tables were created to capture the number of retweets, replies, likes and quotes. Additionally, a two-tailed Mann-Whitney U Test was performed to test the hypothesis that the frequency of interaction with tweets containing the words neurotic and neuroticism-related tweets differs (Nachar, 2008). Lastly, the frequency of tweets mentioning neurotic and neuroticism over time with 30-day intervals was plotted using the Matplotlib library (Hunter, 2007) to capture and compare trends and patterns.

Detection and Comparison of Offensive Language Use (RQ2/H2)

Offensive language detection required pre-processing, such as changing hyperlinks to "http" and mentioned users to "@user" ensuring that they would not impact the sentiment of the tweet's content and conform to the training dataset (Loureiro et al., 2022). However, for the topic modelling prepossessing was not recommended as the replacement led to a deteriorated quality of topics generated by the BERTopic-based model (Heinke, 2023).

Offensive language detection of the roBERTa model was performed to detect offensive tweet content (Barbieri et al., 2020). BERT stands for Bidirectional Encoder Representation from Transformers and thus roBERTa model is a robustly optimized BERT approach (Liu et al., 2019). By employing artificial intelligence to detect the offensive language as well as the topics the human bias was minimized. The model underwent training to detect offensiveness in Twitter data and was used via the TweetNLP API (Camacho-Collados et al., 2022). The model determined each tweet's offensive content and provided a percentage score on which the tweet was categorised as offensive or non-offensive.

An additional chi-square test investigated whether there was a significant difference in the number of offensive tweets between the two keywords (Singhal & Rana, 2015). For this purpose, the overall count of offensive and non-offensive tweets per psychiatric terminology was determined. The effect size was estimated with Cramér's V coefficient (Akoglu, 2018). These two analyses were conducted in Rstudio 2023.06.1 (see Appendix A)

Analysis of Offensive Tweet Content over Time (RQ3/H3a/H3b)

To capture the development of the offensive use of two keywords a graph was plotted with the mean offensive scores. Additionally, the likelihood of offensive content was captured in descriptive statistics covering the overall mean, standard deviation, minimum and maximum for the latter with the according time stamps when they are reached. Furthermore, two visualisations were created to uncover temporal changes in the volume of offensive and non-offensive tweets, one for the keyword neurotic and one for neuroticism.

Lastly, a multiple linear regression analysis with time and keywords as predictors for offensive content was run in Rstudio. The independent variables were time, measured in 30-day intervals, and the keywords neurotic and neuroticism. The dependent variable was the level of offensive content in the tweets. The analysis aimed to determine whether offensiveness rises over time and if the temporal trends significantly differ between the two keywords. To facilitate this, a variable indicating which keyword was mentioned was added to each dataset before combining them into one.

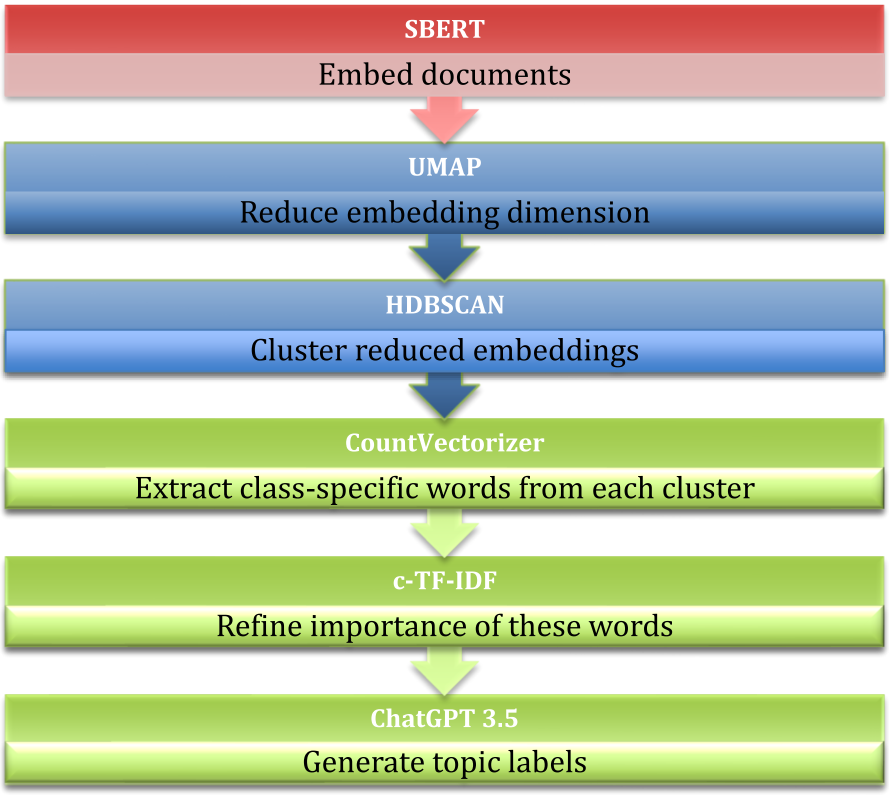
Topic Modelling (RQ4)

The topic modelling with BERTopic assessed the themes discussed in the tweets (Grootendorst, 2022). Topic modelling consists of algorithms that uncover the semantics of texts, and the resulting themes are sorted into topics (Kherwa & Bansal, 2018). Previous models were not tailored to tweets' characteristics, making them less accurate. Therefore, the topic modelling was conducted with BERTopic to overcome the limitations of previous models, in particular concerning tweets (Grootendorst, 2022). The advantages of BERTopic are its flexibility and visualisation options resulting in a more profound analysis of the topics.

To apply the topic modelling on the Twitter data a fitting model was constructed. First, document embeddings were extracted with SBERT, enhancing the topic representation by analysing the complete semantic structure of the tweets (Reimers & Iryna Gurevych, 2019). UMAP helped to decrease the complexity of the embeddings (McInnes et al., 2020). Due to its stochastic nature, repeated running of the model gives varying results, which helps identify the most accurate and comprehensive topics. Subsequently, HBDSCAN, which is a clustering algorithm, grouped the embeddings into clusters based on their semantic similarities (Campello et al., 2013). This step also removed tweets that did not represent relevant themes, ensuring the analysis focused on meaningful data. Then the function CountVectrorizer extracted key candidate words from the embeddings. CountVectorizer converts a collection of text documents to a matrix of token counts, identifying the most frequently occurring words which are likely to represent the main topics (Kristien Margi Suryaningrum, 2023). Next, the class-based term frequency-inverse document frequency (c-TF-IDF) refined the candidate words (Grootendorst, 2022). c-TF-IDF adjusts the importance based on how frequently it appears across different clusters, helping to distinguish more relevant terms from less significant ones. Finally, the pipeline concluded with ChatGPT 3.5 to create final topic names. This customised pipeline is illustrated in Figure 1.

Figure 1

Customised Pipeline for Topic Modelling with BERTopic



Note. The blue-highlighted steps are part of clustering the documents and the green-highlighted steps are part of creating the topic representations

To further benefit from the adaptability of BERTopic an unsupervised topic detection was employed which is a more exploratory approach not using any predetermined classification (Silva et al., 2023). Moreover, Dynamic Topic Modelling provided insights into temporal topic developments hierarchical topic modelling established the interrelation between topics. These techniques added to the interactive and thus engaging and extensive topic visualisations. An intertopic distance map depicting the resemblance of semantic structures and all topics and temporal topic development representations was created.

Python requires the bertopic library to be loaded (Grootendorst, 2022). Although the all-mpnet-base-v2 sentence-transformer required larger computational power it was chosen due to its better results. This model is designed to capture the meaning of sentences and paragraphs making it ideal for clustering and semantic search. In an iterative process, the number of topics was reduced to describe the tweets adequately. To this end, min_topic_size was increased to generate more common topics, and n_neighbours was increased when broadening the overview of topics was necessary. These two parameters were changed in an iterative process to reach a sufficiently broad but clear overview of the topics. Moreover, HDBSCAN automatically combined similar topics based on semantic and thematic resemblance and the CountVectorizer removed stop word topics.

Customised labels with ChatGPT AI were based on the frequently mentioned terms and the context. The frequent key terms were further compared to a list of common insults to identify topics associated with slurs. Visualisations helped to clarify ambiguous topics. Topic trends over time were enabled through nr_bins, which sets the number of intervals the timeframe is divided into.

Results

A total of 430409 tweets were analysed, 427027 contained the term neurotic and 33282 mentioned neuroticism. For simplicity, the tweets mentioning the word neurotic will be referred to as NT and tweets about neuroticism will be referred to as NCMT.

Prevalence of psychiatric terminologies in public discourse (RQ1)

Table 1 captures the descriptive statistics and Mann-Whitney U test results for the tweet-related interaction metrics retrieved by snscrape – retweets, replies, likes and quotes - of NT and NCMT. Among these metrics, like count had the highest average and the greatest variation, reflecting a massive spread in user engagement through likes. NT had a like count mean of 4.38 (SD =141.25). For NCMT, the mean like count was 4.76 (SD = 62.71). All other interaction metrics for both NT and NCMT had means close to zero but differed in their standard deviations. Quotes showed the lowest variance, with a standard deviation of 7.03 for NT and 1.49 for NCMT. Engagement with tweets via replies, retweets, and likes varied increasingly according to the order listed. The means were consistently higher for NCMT than for NT while the standard deviations were higher for NT than for NCMT.

The Mann-Whitney U Test aimed at checking for significant differences in user interaction with the tweets depending on the psychiatric term they contain. The results indicate that there is a significant difference between NT and NCMT for all interaction metrics.

Table 1

Descriptive Statistics and Mann-Whitney U Test Results of the Interaction Metrics of Neurotic- and Neuroticism-mentioning Tweets

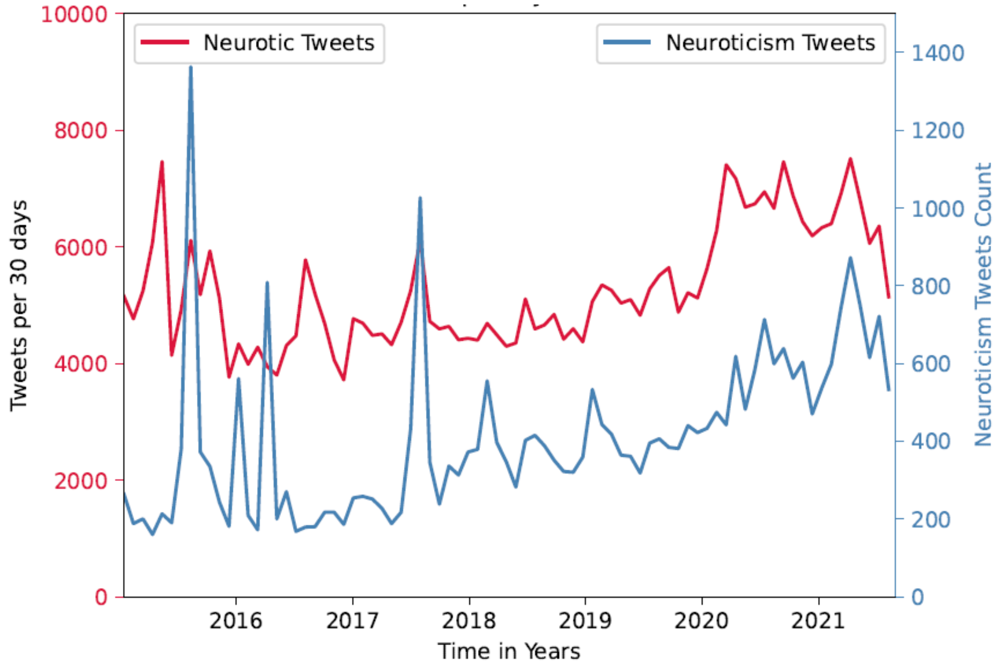
Interaction metrics	Neurotic		Neuroticism		U-statistic	Z-score	p-value
	M	*SD*	*M*	*SD*			
Retweets	0.76	29.22	0.85	10.76	6905470304	-8.6	<.001
Replies	0.48	9.27	0.49	3.09	6921648531	-7.9	<.001
Likes	4.38	141.25	4.76	62.71	7013954197	-3.95	<.001
Quotes	0.09	7.03	0.10	1.49	7011666371.5	-4.05	<.001

As depicted in Figure 2 the frequency of NT consistently outnumbers the posted NCMT between January 2015 and August 2021. Moreover, NT have had a general rising trend over the years with the most notable increases in early 2015 and the beginning of 2020. The increase in early 2015 was possibly linked to the discourse concerning a study about the geographical prevalence of personality traits in Great Britain (Rentfrow et al., 2015). The study was featured in an article in The Guardian, where Sample (2015) highlighted that "Wales was home to a disproportionate number of shy and neurotic people". With the beginning of the Covid-19 pandemic social media, particularly Twitter was used more frequently to discuss the crisis and associated emotions which could explain the heightened number of NTs in early 2020 (Dalili Shoaei & Dastani, 2020). The volume of NT at the start and end of the observation are similar with around 5000 tweets per month. Therefore, the number NT can be considered relatively stable with fluctuations between a little less than 4000 tweets and around 7500 tweets per month.

In contrast, NCMT showed a larger upward with two rapid increases in August 2015 reaching 1300 tweets and in August 2017 reaching above 1000 tweets per month. Nevertheless, the volume of tweets drops immediately after the peaks. The peak in August 2015 may be related to a study that was widely discussed in the media linking creativity to neuroticism (Walten, 2015). The high number of NCMT in 2017 could be explained by the viral scandal of a Google employee claiming that women are less suitable to work in tech due to their high levels of neuroticism (Chuck, 2017).

Figure 2

Frequency of Neurotic- and Neuroticism-related Tweets from January 2015 to August 2021



Offensive Public Use of Psychiatric Terminology (RQ2/H1)

Based on the offensive language detection 31% of NT and 8% of NCMT were labelled as offensive. To investigate whether the relationship between the mentioned psychiatric terminology and the offensive nature of the tweet is indeed significant a chi-square test of independence was conducted. First, the number of offensive and non-offensive tweets were captured for NT and NCMT as seen in Table 2. Second, the chi-square test was run using this 2x2 table resulting in a χ^2 (1, N = 460309) = 7694.6, p < .001 which shows that NT were significantly more likely to be offensive than NCMT.

Table 2

Number of offensive and non-offensive tweets per keyword

Tweet type	Neurotic	Neuroticism
Not offensive	295403	30577
Offensive	131624	2705

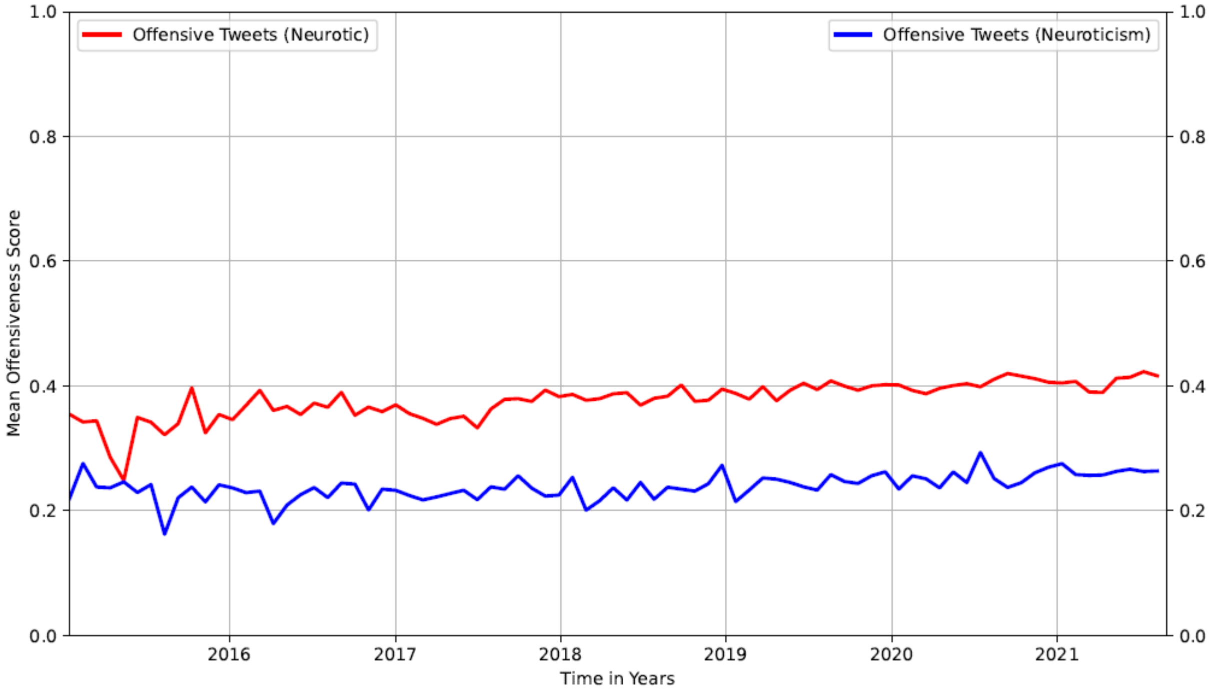
Additionally, Cramér's V was determined to assess the effect size of the relation between terminology and offensiveness. A Cramér's V of 0.13 [confidence interval: 0.12, 1.00] was detected, suggesting that although the use of neurotic is associated with a more offensive nature of tweets, the strength of this association is weak.

Offensive content over time (RQ3/H3a/H3b)

The likelihood of offensive content in tweets from 2015 up to 2021 is displayed in Figure 3 and variance measures are supplementally captured in Table B1. The red line represents the course of NT with a tendency to be offensive and the blue line for NCMT. Both lines fluctuate over the years. However, the likelihood of offensive content in NCMT is more stable with a standard deviation of 0.02 compared to the NT that have a standard deviation of 0.03. The average likelihood for offensive content of NT varies between a minimum of 0.25 in May 2015 and a maximum of 0.42 in July 2021. For NCMT the mean likelihood reaches a minimum of 0.16 in August 2015 and a maximum of 0.29 in July 2020. The overall mean likelihood of offensive NT is 0.38 and exceeds the 0.23 of NCMT. This difference can also be seen as a general tendency over the years in Figure 3 with an exception in 2015 when the likelihood of offensive NT dropped and became equal to the one for offensive NCMT. Overall, these findings suggest that the proportion of offensive tweets using NT and NCMT is relatively stable over time.

Figure 3

Mean Offensiveness Scores of Neurotic- and Neuroticism-mentioning Tweets from January 2015 to August 2021



Figures 4 and 5 depict the volume of tweets differentiated into offensive and not offensive over the collected time. Figure 4 displaying NT shows moderate fluctuations for offensive tweets. The offensive use of the term neurotic at the beginning of the observation was below 2000 tweets per 30 days and raised to above 2000 tweets per 30 days throughout the observation. This increase may be related to the overall increase in the number of NT displayed in Figure 2 highlighting the steady likelihood for offensive content as illustrated in Figure 3. In 2015 more not-offensive tweets were posted but from mid-2021 there were more or equal amounts of offensive tweets. Not-offensive tweets peaked once at the beginning of 2015 rising from a little above 3000 to above 5000 but then dropped rapidly below 3000.

Figure 4

Offensive and Non-offensive Neurotic-mentioning Tweets Volume over Time

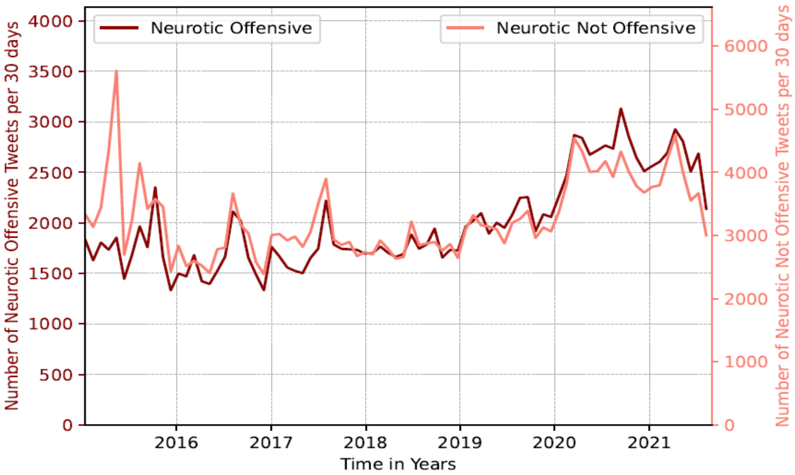
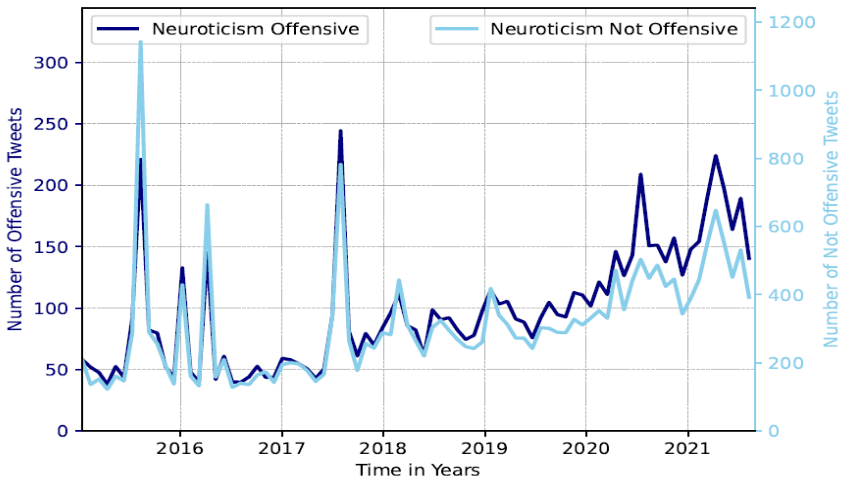


Figure 5 depicts a growing trend of offensive and not-offensive NCMT. The highest number of not-offensive NCMT is depicted in mid-2015 with almost 1200 tweets. Another notable increase is in mid-2017 with almost 800 tweets. These peaks are consistent with the ones observed in Figure 2. Relatively speaking, the peak in 2015 was more pronounced for not-offensive NCMT whereas the second in 2017 was more notable for the offensive tweets. Additionally, the offensive tweets had a relatively larger growth tendency from 2020 onwards. The offensive and not-offensive tweets behave similarly over time highlighting that the increases and decreases are closely related to the overall trend of neuroticism-related trends.

Figure 5

Offensive and Non-offensive Neuroticism-Mentioning Tweets Volume over Time



Given that the overall proportion of offensive NT and NCMT is relatively stable over time, it is not surprising that Figures 4 and 5 show similar behaviour to Figure 1. This stability indicates that despite fluctuations in the volume of tweets, the ratio of offensive to non-offensive content remains fairly consistent. The consistent behaviour suggests that the dynamics observed of offensive and non-offensive tweets are influenced by similar underlying factors over the observed period, making observed patterns expected.

A multiple regression analysis was run to test two hypotheses: first, whether keywords neurotic and neuroticism are used more offensively with increasing time, and second, whether the offensive use of each term develops differently. Both hypotheses were supported by the analysis examining the predictive power of the mentioned keyword and time interval on offensive content. It revealed a significant model, $F(3,460305) = 360.00$, $p < .001$ with $R^2 = .02$, indicating that approximately 2% of the variance in offensiveness can be explained by the predictors.

According to the regression coefficients each of the predictors was significant (Table 3). More precisely, the keyword variable significantly predicted offensiveness (B = -0.20, SE = 0.006, $t = -36.97$, $p < .001$), showing that the use of neuroticism is associated with a decrease of offensiveness by 0.203 units compared to neurotic when controlling for time. For the second significant predictor time (B = 0.002, SE = 0.00003, $t = 53.40$, $p < .001$) the results indicate that each 30 days offensiveness rises by 0,0015 units when the used keyword remains the same. Lastly, the interaction between time and keyword was significant (B = -0.001, SE = 0.0001, t = -5.91, p < .001), suggesting that with time progressing the effect of keyword on offensiveness diminishes.

Table 3

Multiple Regression Analysis of Mentioned Keyword and Time on Offensiveness

Variable	*B*	*SE*	*t*	*p*	95% CI
Keyword	-0.20	0.006	-36.97	<.001	[-0.21, -0.19]
Time	0.002	0.00003	53.40	<.001	[0.0015, 0.0016]
Keyword x Time	-0.001	0.0001	-5.91	<.001	[-0.0008, -0.0004]

Note. Keyword indicated whether neurotic or neuroticism was mentioned, and time indicated the number of 30-day intervals since the start of the observation period.

Identified Topics (RQ4)

Topic modelling was employed via BERTopic to identify the most frequently discussed topics of NT and NCMT. A total of 71 topics were extracted from NCMT and 89 topics from NT. Tables containing a complete list of all topic labels, their count, their portion, their keywords and example tweets per psychiatric term can be accessed with the links in Appendix A. Moreover, all links to the interactive visualisation are provided (see Appendix A).

Tweets which notably deviated from the established topics were categorised as outliers. The NT have a higher proportion of outliers with them making up 49,79% whereas for NCMT it is only 34,98%. Moreover, the NT as a terminology have topics containing between 310 to 66476 tweets. For neuroticism, the smallest topic contains 21 tweets and the largest 16457. For both groups of tweets, the largest topic is discussed far more often than any other topic.

Key Topics of Tweets Containing the Term Neurotic

The top five topics of NT are titled*: 'melodramatic personality types',' pet behaviour study'*, *'Trump's chaotic presidency'*, *'musical outsider exploration'* and *'writer's creative struggles'*. *'Melodramatic personality types'* is the most frequently discovered topic covering 15.57% of NT using the word neurotic. The ten most frequent keywords of the topic reveal that the topic is related to the lyrics of the song 'Basket Case' from Green Day (DeLoye, 2024). The tweets talk about the following extract: "I am one of those melodramatic fools. Neurotic to the bone, no doubt about it" (see Appendix A). The keywords *like* and *love* appear frequently in the tweets linked to *Melodramatic personality types'*. These words are not part of the lyrics but rather reflect the users' emotional reactions to the song. The bar chart of the commonly used terms associated with 'melodramatic personality types' shows that *she's* is mentioned second most (see Appendix A).

'Pet behaviour study' covers 4.01% of the tweets and deals with different studies that examine pets and their owner's behaviour neurotic. One of the main debates was about the differences between *cats* and *dogs* and who prefers which pet (Adomaite & Akavickaitė, 2021). However other studies checked for correlations between neurotic tendencies in pet owner and their pets (Aada Ståhl et al., 2023). In addition to study-related content, the *'Pet behaviour study'* contains tweets where users express their *love* towards their pets.

The third topic with 2.56% of tweets is '*Trump's chaotic presidency*' consisting of related reactions. *Trump* was known for using Twitter for attacks on political opponents like the Democratic National Committee as well as *Mika* Brzezinski and *Nancy* Pelosi (Diamond

& Collinson, 2016; Stableford, 2020). Many of his Twitter disputes characterised by name-calling and insults were covered by *CNN* (Winberg, 2017). Apart from his real last name *Trump,* there were frequent references to his social media name *realdonaldtrump* (see Appendix A). Tweets on the '*Trump's chaotic presidency*' topic often discuss Tump's Twitter activities and disputes with other politicians.

The '*Musical outsiders exploration'* topic covered 2.41% of tweets. In this topic, people predominantly shared their enthusiasm for the supergroup Neurotic *Outsiders* consisting of members of other known bands, for example, Steve Jones from the Sex Pistols and Duff McKagan from Guns N' Roses. Many tweets include links to music videos or streams of the band, illustrated by the use of keywords such as *YouTube*, *song*, *listen*, and *new* (see Appendix A).

The fifth topic is the '*writer's creative struggles'* constituting 1.795% of the tweets. Based on the most frequent terms of this topic it primarily discusses the *writing* of *books* (see Appendix A). More precisely, in this topic users share challenging experiences as *writers* with narratives or cite quotes from books whose titles include the term neurotic.

Overall, it appears that the term neurotic is mostly used in non-derogatory contexts, with the notable exception of the political discourse regarding Donald Trump which is related to incivility and derogatory remarks.

Key Topics of Tweets Containing the Term Neuroticism

For neuroticism, the top five topics are '*personality traits analysis'*, *'unique man descriptions'*, *'cat owner dynamics'*, *'spiritual belief and culture'* as well as *'Jewish identity traits'*. 'Personality traits analysis' is the largest topic constituting 49,45% of the tweets. It explores the big five personality traits, their manifestation, and their associations. The keywords of this topic include the other dimensions of the big five personality traits supporting that the model is being referred to (see Appendix A). Furthermore, the terms *high* and *higher* suggest that correlations are being discussed mainly between neuroticism and the also mentioned *depression* and *creativity*. The scientific focus of this largest topic can be attributed to neuroticism being generally the most researched personality trait (Cassiello-Robbins et al., 2017).

The second topic namely, *'unique man descriptions'*, applies to 1.91% of the tweets and includes expectations for a *perfect* man. In many tweets, public figures and fictional characters are referred to who represent the optimal level of traits such as neuroticism thus the high number of male names among frequently used key terms.

'Cat owner dynamics' means the relationship between owning a *pet*, particularly *cats*, and the neurotic feature of the owner discussed in 0.84% of cases. Similar to the *'Pet behaviour study*' topic from NT this topic is dominated by research revealing cat owners scoring higher on neuroticism than *dog* owners and how this neuroticism affects the pets. In the context of neuroticism, Twitter users posted studies exploring cat personality traits and how they differ between domestic cats and lions (Gartner et al., 2014).

The fourth topic is connecting '*spiritual belief and culture'* to neuroticism which can be found in 0.76% of the tweets. According to the keywords the topic mainly deals with *religion* and its correlation with neuroticism. Again, tweets are tied to psychological studies concerning associations between spiritual *beliefs* and *cleanliness* (Fetterman, 2016). Furthermore, the low neuroticism among catholic priests is mentioned (Cerasa et al., 2016).

Furthermore, the fifth topic with 0.65% of the NCMT relates specifically to the scientifically not supported *Jewish* stereotype regarding high neuroticism (Cosman, 2018; Jacobs, 2008). The portrayal of this stereotype in the entertainment industry is discussed. The most prominent keywords highlight the repeated sharing of an article titled: "An ode to 'Girls' *Lena Dunham* and modern Jewish Neuroticism" from the *Times of Israel* is shared multiple times making the title part of the most prominent keywords (Friedman, 2017). Moreover, some tweets link the supposed neuroticism of *Jews* to the history of the Holocaust.

Likewise, it seems that among these top five topics, neuroticism was mostly used for non-derogatory discussions in four out of five cases, with Jewish stereotyping being the exception.

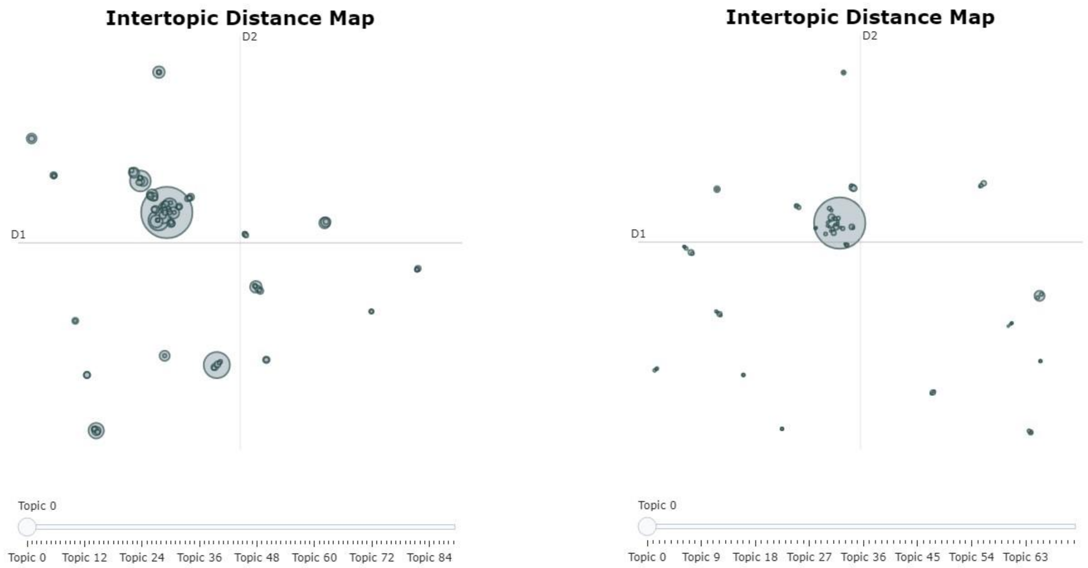
Topics with Insulting Terms and Clusters

An additional search for insults among the topic terms revealed that no common insults were found in any of the NCMT topics. However, insults in three NT topics were detected. *'Melodramatic personality types'* included the term *fools* as it is part of the song's lyrics frequently shared within this topic. The top sixth topic '*emotional acronym explanation'* included multiple derogatory namely *fucked*, *freaked* and *fuckedup*. Lastly, the top 27th topic of the NT called '*truthful politician's message'* entailed the word *pigheaded*.

Each topic is presented as a dot varying in size depending on the number of associated tweets and placed in relation to the other topics. The left figure shows one very large cluster of topics and multiple others that stand out. In comparison, the right figure depicts one central cluster which exceeds the size of all other topics as well as clusters.

Figure 6

Intertopic Distance Map for Neurotic-related (left) and Neuroticism-related Tweets (right)



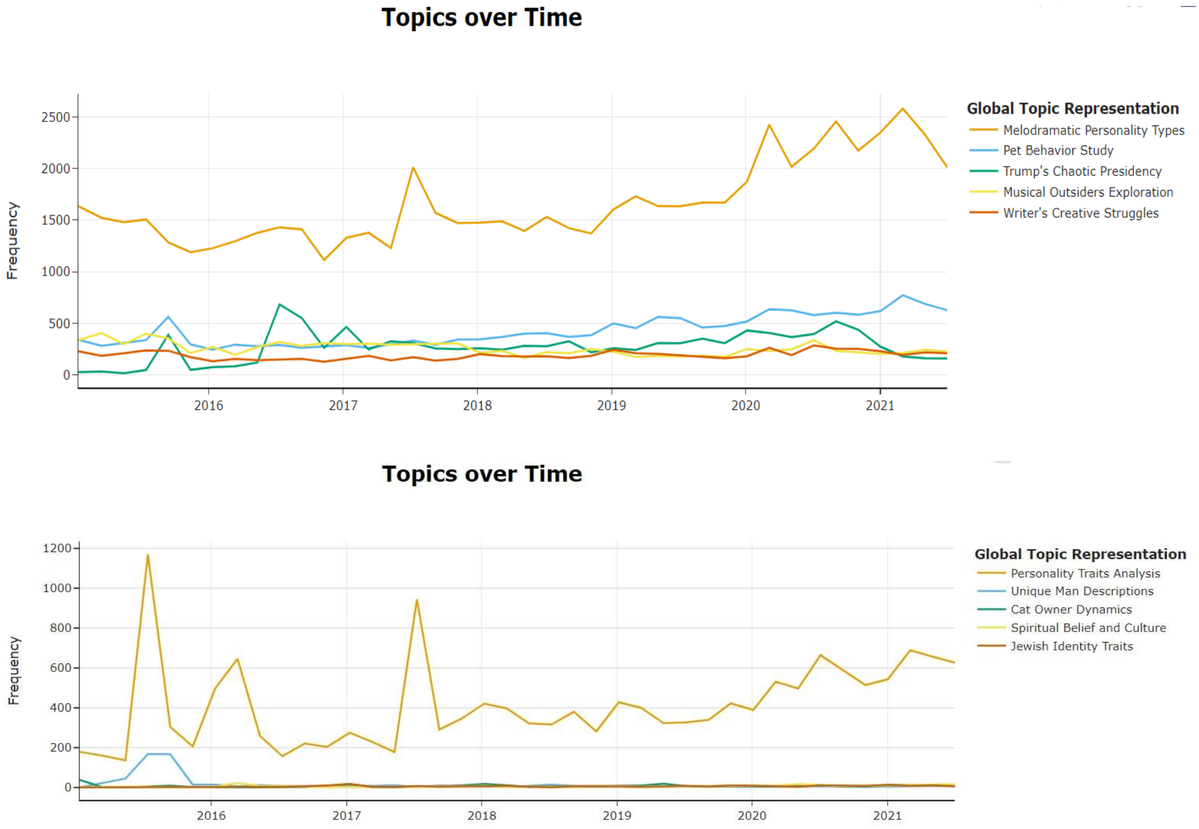
Topics over Time

The topic development over time shown in Figure 7 illustrates which topics were more prominently discussed than others. The discourse related to '*melodramatic personality types'* for NT and *'personality trait analysis'* for NCMT strikingly exceeds all other topics over the entire time suggesting they are of the highest interest to the public. '*Pet behaviour* study' is displayed with a lower frequency and only minor fluctuations reflecting a steady interest whereas peaks in 2016 and 2017 indicate increased discussions about *' Trump's chaotic presidency'*. The topics *'Musical outsiders exploration'* and *'writer's creative struggles'* are consistent over time however with low frequencies. For the NCMT, next to the largest topic only the 'u*nique man descriptions'* topic rose in public interest temporarily in 2015. The graph highlights that besides the top 1 topic, the tweets of the rest topics are barely shared.

Additionally, fluctuations display the varying extent of topic discussions. Furthermore, Figure 7 reflects trends that were already observed in Figure 3. For NT, the peaks from 2020 onwards were defined by heightened discourse concerning '*melodramatic personality types'*. All increased frequencies in NCMT can be attributed to '*personality trait analysis'* and in 2015 as well to debates regarding the '*unique man descriptions'*.

Figure 7

The Top Five Topics over Time for Tweets containing the terms neurotic (upper) and neuroticism (down)



Discussion

Summary of Key Findings

The purpose of this study was to gain a scientific understanding of the public use of psychiatric terminologies, namely neurotic and neuroticism and how their usage has evolved from January 2015 to August 2021. Consequently, four research questions and hypotheses were examined.

The first key finding was the predominance of the word neurotic in the common parlance compared to neuroticism (RQ1). The results support the hypothesis that interactions with tweets differ between NT and NCMT (H1). Interactions with NCMT were more frequent on average, whereas interactions with NT varied more widely. Heightened use of the word neurotic was observed in early 2015 driven by a study on the prevalence of personality traits in Great Britain, and again in early 2020, as discussions related to the Covid-19 pandemic intensified. The term neuroticism saw increased mentions in August 2015 due to a published study linking neuroticism to creativity, and in August 2017 following a scandal of a Google employee's controversial comment about women's suitability for tech jobs based on their perceived high levels of neuroticism.

Moreover, the study revealed higher denigrative use of the term neurotic 31% of the time compared to neuroticism which was used offensively only 8% of the time (RQ2/H2). Both keywords predicted an increase in offensive use over time (RQ3/H3a). The polarising content of NCMT was more stable, with notable rises in mid-2015 and mid-2017 and a more prominent increase from 2020 onwards, likely due to the same factors of increased discussion mentioned earlier. For NT non-offensive tweets showed more fluctuations than offensive ones. Lower levels of offensiveness were consistently associated with neuroticism (H3b).

Lastly, the topics in which the words neurotic and neuroticism were used were identified (RQ4). The term neurotic appeared in 89 topics, encompassing more tweets than neuroticism which was found in 71 topics. Most common themes including the word neurotic, in descending order, were *'melodramatic personality types', 'pet behaviour study', 'Trump's chaotic presidency', 'musical outsiders exploration'* and *'writer's creative struggles'*. For neuroticism themes included: '*personality traits analysis', 'unique man descriptions', 'cat owner dynamics', 'spiritual belief and culture'* and *'Jewish identity traits'*. The dominant topic was *'melodramatic personality types'* for the NT, which was interpreted as related to the lyrics of the song 'Basket Case' by Green Day. For the NCMT, the topic '*personality traits analysis'* was most commonly seen, and was interpreted to be related to the exploration of the big five personality traits, their manifestations, and their associations. Spikes in interest for the term neurotic were observed between 2016 and 2017 regarding '*Trump's chaotic presidency'*, while for neuroticism discussions of 'unique man descriptions' were trending in 2015.

Links to Previous Research and Interpretation

Previous research documented that teenagers often describe individuals with mental health struggles using derogatory references, a phenomenon evident in different types of mass media. (Rose et al., 2007). The use of the label neurotic is consistent with these findings, illustrating that the derogatory language spreads onto social media and over different age groups, reflecting a general tendency within public discourse. The high prevalence of the keyword neurotic in colloquial language indicates a higher familiarity with the term, but possibly insufficient understanding of its meaning leading to stigma. Neurotic is the more common term, likely due to its introduction with neurosis in 1769, whereas the concept of neuroticism arose in 1947 (Flehmig et al., 2007; Pinero, 1983; Sims, 1983). Still, out of the five most common topics for NT and NTCM, only two were interpreted to be related to derogatory use: '*Trump's Chaotic presidency'* and '*Jewish identity traits'*. This indicates that the majority of discourse involving these terms was neutral or non-offensive rather than derogatory.

In terms of engagement, tweets containing the term neurotic had, on average, fewer interactions compared to those containing neuroticism. This is consistent with findings that non-toxic tweets generally receive higher engagement (Salehabadi et al., 2020). Although, toxicity, including derogatory language such as the label neurotic, affects interactions, highly toxic tweeds tend to generate more user interaction than less but still toxic ones. This variability in interaction can be attributed to the increased derogatory content, as seen with tweets containing the term neurotic. Nonetheless, no definitive explanation exists for why offensive language use does not show higher overall engagement. A study by Guenther et al. (2023) explains that engagement with scholarly tweets depends on content-related factors such as emotions, humour or discussion elements and functional factors such as hashtags, mentions, links or videos. Therefore, the presence of a single term in a tweet is not sufficiently meaningful to predict interactions, a combination with other engaging features could explain the different levels of user interaction.

The observed peaks in shared NT and NCMT often followed single events or published studies. However, these temporary spikes should be differentiated from long-term engagement trends, which are better revealed through comprehensive topic analysis. This differentiation was revealed in the topic analysis, which highlighted that while certain events led to temporary spikes, consistent engagement over time indicated deeper trends and public interest. Similarly, the persistence of trends in social media, as noted by Asur et. al (2021), shows that while most topics disperse quickly, those with resonating content exhibit long-term engagement.

The higher frequency of offensiveness in NT compared to NCMT indicates a concerning trend in public discourse. According to Antonci et al. (2019), online incivility is growing demonstrated by the higher prevalence of the term neurotic and its frequent offensive use. This derogatory use can undermine efforts to destigmatize mental health issues, by reinforcing negative stereotypes and stigmatizing individuals who might exhibit neurotic traits. Past research showed similar stigmatizing misuse of psychiatric terms, particularly the term schizophrenia (Moriceau et al., 2022). One survey analysed the public use of schizophrenia and schizophrenic as one without comparing them (Observatoire de la Société et de la Consommation (L'Obsoco), 2015). As indicated by Padilla Cruz (2019) likelihood of offensiveness differs between adjectives and nouns. The results highlight the greater potential of adjectives to be used offensively. Hence, misuse of psychiatric terminologies varies, with some terminologies being closer related to stigma than others.

Concerning temporal trends, previous research has not shown significant trends within stigmatizing language use (Pavlova & Berkers, 2020b). However, the study of Pavlova and Berkers (2020b) did not consider the offensive notion of the terms neurotic and neuroticism. Lacking scientific examination of the terms possibly reduced their coverage in anti-stigma campaigns which usually focused on schizophrenia or depression (Walsh & Foster, 2021). The growing offensive language use is consistent with the trend of online incivility linked to heated political discussions and specific events like the COVID-19 pandemic and the Trump presidency implying an influence of broader societal contexts on the use of psychiatric terms (Antonci et al.,2019; Theocharis et al., 2020). The rise in offensiveness reflects growing political polarisation, the disinhibition effect online and the normalisation of toxic language (Antoci et al., 2019; Esfahani et al., 2017; Yarchi et al., 2020).

The identified topics demonstrate findings from previous literature. For instance, topics dealing with '*musical outsiders exploration',* '*Trump's chaotic presidency'* and *'Jewish identity traits'* reflect popular Twitter themes such as music, news, politics, and religion (Lee et al., 2011; Zucker, 2016). Research about mental health discussions however reveals other themes for instance education/research/schooling, problematization, feelings, community/awareness/events and anti-stigma (Pavlova & Berkers, 2020a). While this study did not reveal any topics specifically related to the last four themes, it did highlight important themes of NCMT, which were often related to research. For instance, the topic *'personality traits analysis'* focuses on the big five personality traits, implying a more scientific notion, while the topics '*Trump's chaotic presidency'* and '*melodramatic personality types'* create the impression that the term neurotic is used more casually. Interestingly, discussions about Trump were tied to the context of mental health before (Pavlova & Berkers, 2020a). Trump being present in the topics of both keywords stresses his significant role in perpetuating stigma (Goodman, 2023). The term neurotic was mentioned as the name of a book or a band as well as part of Green Day's song lyrics name of suggesting an influence of pop culture on common language use. The often-posted song's lyric "I am one of those melodramatic fools. Neurotic to the bone, no doubt about it" appears self-stigmatizing (DeLoye, 2024). The topic 'm*elodramatic personality types'* also indicated discourse about females with neurotic tendencies illustrated by the frequent mention of *she's*. The stereotype of the neurotic female is congruent with previous research describing the former belief of hysterical neurosis being an exclusive female diagnosis (McGrory, 1980; Tasca et al., 2012). The presence of other insults mentioned in the discourse involving the keyword neurotic underscores its association with derogatory language and its negative connotations in public discourse.

Strengths, Limitations and Future Research

Certain limitations of this study could be addressed in future research for example the scraping of tweets should be refined as many tweets were included based on the keyword being part of names or lyrics. This inclusion could misrepresent how the terms are usually used in public discourse. Consequently, new studies should adapt their scraping to collect only data relevant to the studies' aim. A specific filter recognising the use of keywords as part of names or quotes is required. Nonetheless, tweets with indirect use of keywords generated insights into potential influence on language use which can be valuable depending on the research aim. Additionally, the scraped data allowed for longitudinal analysis up to 2021, omitting any recent trends that may be of importance. Therefore, research on the use and misuse of psychiatric terminologies should be continued to detect any prospective developments requiring attention and countermeasures.

The present study stands out for the application of artificial intelligence-driven topic modelling and offensive language use which is a new approach to eliminating human-made errors and biases and allows for more sophisticated but less time-consuming processing of large quantities of data. Nonetheless, the mechanism of AI, especially unsupervised learning methods, may introduce unperceivable biases or subjectivity. These limitations should be acknowledged when analysing the results.

Moreover, the generalizability of results must be considered. Although the present results clearly support that the term neurotic predicts more frequently offensive tweet content, it is important to recognize potential limitations due to the chosen platform and language. Although Twitter offers easy access to rich observations compared to other platforms, the findings are not representative of all social media users or face-to-face communication. However, analysis of online communication generates insights into otherwise hidden preconceptions of the public as it depicts a more unfiltered picture of opinions due to social media's anonymous nature. Future studies could dive further into the use of the terms in different languages and investigate face-to-face communication of terms such as neurotic and neuroticism with the use of surveys.

Another area for future research could be the exploration of users posting offensive tweets. Analysing their demographics and characteristics could provide insights into the source of stigma. By understanding who is more likely to use derogatory language, targeted interventions could be developed to address these specific groups. This research could include examining the influence of user anonymity, gender, age, and social media behaviour patterns.

Practical Implications

Despite these limitations, these results suggest several practical implications. For instance, educational interventions should be developed to promote accurate and respectful use of psychiatric terminology in public discourse. The content of these interventions should address prejudice about people with neurotic tendencies as presented in this study by discussing the impact of stigmatizing language and the current scientific indications for the use of the term neurotic detached from its history. Already existing mental health awareness campaigns aiming to reduce stigma should expand their focus beyond depression and schizophrenia and consider the offensive use of the term neurotic.

Second, the recorded incidents of misusing the term neurotic can inform platform policies to review their offensive language detection and stimulate new moderating measures that provide feedback to users. These measures can help to spread awareness and reinforce civilised discourse on social media.

Conclusion

This study highlights the prevalent misuse and stigmatizing connotations linked to the term neurotic in public online discourse, compared to the scientifically dominated term neuroticism. The findings indicate an insufficient mental health literacy and comprehension of the appropriate use of psychiatric terminologies, fuelling stigma towards neurotic personality traits. Online platforms, particularly Twitter, appear to uphold and reinforce the denigrative use of terms and thus undermine efforts to destigmatize mental health issues. Moreover, factors such as political polarisation and the normalisation of toxic language are possible explanations for the sustained offensive use of psychiatric terms. Nevertheless, the study also shows research-based discussions around neuroticism indicating a certain level of appropriate use. Interventions such as educational interventions and stricter policies are recommended to counteract the offensive use and spread awareness. However, continuous research monitoring the ongoing development of language use and misuse is necessary for such interventions. In this regard, artificial intelligence can be useful while recognising the limited algorithmic transparency. Overall, the study points out the need for further action in terms of de-stigmatising efforts regarding public attitudes requiring collaboration between researchers, policymakers, and educators.

References

Aada Ståhl, Salonen, M., Hakanen, E., Mikkola, S., Sini Sulkama, Lahti, J., & Hannes Lohi. (2023). Pet and Owner Personality and Mental Wellbeing Associate with Attachment to Cats and Dogs. *IScience*, *26*(12), 108423–108423. https://doi.org/10.1016/j.isci.2023.108423

Adomaite, L., & Akavickaitė, A. (2021, March 22). *40 Of The Best "Dog People Vs. Cat People" Tweets*. Bored Panda. https://www.boredpanda.com/cat-vs-dog-people-tweets/

Akoglu, H. (2018). User's Guide to Correlation Coefficients. *Turkish Journal of Emergency Medicine, 18*(3), 91–93. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107969/

Alshaabi, T., Dewhurst, D. R., Minot, J. R., Arnold, M. V., Adams, J. L., Danforth, C. M., & Dodds, P. S. (2021). The growing amplification of social media: measuring temporal and social contagion dynamics for over 150 languages on Twitter for 2009–2020. *EPJ Data Science*, *10*(1). https://doi.org/10.1140/epjds/s13688-021-00271-0

Antoci, A., Bonelli, L., Paglieri, F., Reggiani, T., & Sabatini, F. (2019). Civility and trust in social media. *Journal of Economic Behavior & Organization*, *160,* 83–99. https://doi.org/10.1016/j.jebo.2019.02.026

Arboleda-Flórez, J., & Stuart, H. (2012). From Sin to Science: Fighting the Stigmatization of Mental Illnesses. *The Canadian Journal of Psychiatry, 57*(8), 457–463. https://doi.org/10.1177/070674371205700803

Asur, S., A. Huberman, B., Szabo, G., & Wang, C. (2021). Trends in Social Media: Persistence and Decay. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), 434–437. https://doi.org/10.1609/icwsm.v5i1.14167

Bączkowska, A. (2021). "You're too thick to change the station" – Impoliteness, insults and responses to insults on Twitter. *Topics in Linguistics*, *22*(2), 62–84. https://doi.org/10.2478/topling-2021-0011

Badcock, C. (1992). *Essential Freud: An Introduction to Classical Psychoanalysis (Second Ed.)*. https://doi.org/10.13140/RG.2.1.3874.3129

Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). *TWEETEVAL: Unified Benchmark and Comparative Evaluation for Tweet Classification*.

Benesch, S. (2014). Defining and diminishing hate speech. *State of the World's Minorities and Indigenous Peoples, 2014,* 18–25.

Bisiada, M. (2022). Discourse and Social Cohesion in and After the Covid-19 Pandemic. *Media and Communication*, *10*(2), 204–213. https://doi.org/10.17645/mac.v10i2.5150

Britannica. (2019). neurosis | Definition, Types, Treatment, & Facts | Britannica. In *Encyclopædia Britannica*. https://www.britannica.com/science/neurosis

Cai, Y., & Sun, J.-T. (2009). Text Mining. In *Encyclopedia of Database Systems* (pp. 3061–3065). https://doi.org/10.1007/978-0-387-39940-9_418

Camacho-Collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., Boisson, J., Espinosa-Anke, L., Liu, F., Martínez-Cámara, E., Medina, G., Buhrmann, T., Neves, L., & Barbieri, F. (2022). TweetNLP: Cutting-Edge Natural Language Processing for Social Media. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2206.14774

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. *Advances in Knowledge Discovery and Data Mining*, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

Carnaghi, A., & Bianchi, M. (2017). Derogatory Group Labeling. *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.435

Cassiello-Robbins, C., Wilner, J. G., & Sauer-Zavala, S. (2017). Neuroticism. *Encyclopedia of Personality and Individual Differences*, 1–6. https://doi.org/10.1007/978-3-319-28099-8_1256-1

Cerasa, A., Lombardo, G., Tripodi, D., Stillitano, E., Sarica, A., Gramigna, V., Martino, I., Pullera, A., Tigani, S., De Carlo, Y., Idone, M., Scaglione, A., Ziarelli, E., Vasta, R., Donzuso, G., Rizzo, M., & Zucaro, D. L. (2016). Five-factor personality traits in priests. *Personality and Individual Differences*, *95*, 89–94. https://doi.org/10.1016/j.paid.2016.02.040

Chowdhary, K. R. (2020). Natural Language Processing. *Fundamentals of Artificial Intelligence*, 603–649. https://doi.org/10.1007/978-81-322-3972-7_19

Chuck, E. (2017, August 7). Google Employee’s Anti-Diversity Manifesto on Women’s “Neuroticism” Goes Viral. *NBC NEWS*. https://www.nbcnews.com/business/business-news/google-employee-s-anti-diversity-manifesto-women-s-neuroticism-goes-n790401

Chung, M. (2019, March 23). *What it Means to Be “Neurotic” (Clinically reviewed by Cynthia V. Catchings, Clinical Therapist)*. Talkspace. https://www.talkspace.com/blog/what-it-means-to-be-neurotic/

Competiello, S. K., Bizer, G. Y., & D. Catherine Walker. (2023). The Power of Social Media: Stigmatizing Content Affects Perceptions of Mental Health Care. *Social Media and Society*, *9*(4). https://doi.org/10.1177/20563051231207847

Corrigan, P. W., & Watson, A. C. (2002). Understanding the impact of stigma on people with mental illness. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *1*(1), 16–20.

Cosman, S. (2018). Jewish religious confrontation and neuroticism. *Journal of Research in Social Psychology. Faculty of Social Sciences | University of Buenos Aires (UBA)*, *4*(1).

Dalili Shoaei, M., & Dastani, M. (2020). The Role of Twitter During the COVID-19 Crisis: A Systematic Literature Review. *Acta Informatica Pragensia*, *9*. https://doi.org/10.18267/j.aip.138

Damanik, T. E. S. (2023). Examining Freedom of Speech on Twitter from an International Law Perspective. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4122003

DeLoye, L. (2024, February 9). *Billie Joe Armstrong Reveals The "Best Decision" He's Made As A Songwriter*. 100.7 WZLX. https://wzlx.iheart.com/content/2024-02-08-billie-joe-armstrong-reveals-the-best-decision-hes-made-as-a-songwriter/

Devlin, J., Chang, M.-W., Lee, K., Google, K., & Language, A. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Diamond, J., & Collinson, S. (2016, July 27). *Dems accuse Trump of disloyalty over Clinton emails | CNN Politics*. CNN. https://edition.cnn.com/2016/07/27/politics/donald-trump-vladimir-putin-hack-hillary-clinton/index.html

Elm, E. von, Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*, *335*(7624), 806–808. https://doi.org/10.1136/bmj.39335.541782.ad

Esfahani, P. M., Sutter, T., Kuhn, D., & Lygeros, J. (2017). From Infinite to Finite Programs: Explicit Error Bounds with Applications to Approximate Dynamic Programming. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1701.06379

Fernandez, D. K., Singh, S., Deane, F. P., & Vella, S. A. (2022). Exploring Continuum and Categorical Conceptualisations of Mental Health and Mental Illness on Australian Websites: A Systematic Review and Content Analysis. *Community Mental Health Journal*. https://doi.org/10.1007/s10597-022-01005-w

Fetterman, A. K. (2016). On God-Belief and Feeling Clean. *Social Psychological and Personality Science*, *7*(6), 552–559. https://doi.org/10.1177/1948550616641474

Flehmig, H. C., Steinborn, M., Langner, R., & Westhoff, K. (2007). Neuroticism and the mental noise hypothesis: Relationships to lapses of attention and slips of action in everyday life. *Psychology Science*, *49*(4), 343–360.

Friedman, G. (2017, February 11). *An ode to "Girls," Lena Dunham and modern Jewish neuroticism*. The Times of Israel. https://www.timesofisrael.com/an-ode-to-girls-lena-dunham-and-modern-jewish-neuroticism/

Gartner, M. C., Powell, D. M., & Weiss, A. (2014). Personality structure in the domestic cat (Felis silvestris catus), Scottish wildcat (Felis silvestris grampia), clouded leopard (Neofelis nebulosa), snow leopard (Panthera uncia), and African lion (Panthera leo): A comparative study. *Journal of Comparative Psychology*, *128*(4), 414–426. https://doi.org/10.1037/a0037104

Gillette, H. (2022, March 25). *Neuroticism: What It Means, Signs, and Tips to Cope (Medically reviewed by Dr. Bethany Juby, Clinical Psychologist)*. Psych Central. https://psychcentral.com/health/all-about-neuroticism

Goodman, M. (2023). *Trump and the Performance of "The Other" and Stigmatization: The Twitterization of Newsmaking and Monetizing*. 197–222. https://doi.org/10.1007/978-3-031-36099-2_9

Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. *European Economic Review*, *136*, 103772. https://doi.org/10.1016/j.euroecorev.2021.103772

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2203.05794

Guenther, L., Wilhelm, C., Oschatz, C., & Brück, J. (2023). Science communication on Twitter: Measuring indicators of engagement and their links to user interaction in communication scholars' Tweet content. *Public Understanding of Science*, *32*(7), 860–869. https://doi.org/10.1177/09636625231166552

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text Mining in Big Data Analytics. *Big Data and Cognitive Computing*, *4*(1), 1. https://doi.org/10.3390/bdcc4010001

Heinke, J. (2023). *"You are what you Tweet!" -Analysing Pro-Anorexia and Pro-Recovery Messages On Twitter Using Transformer-Based Text Mining Applications*.

Henderson, C., Evans-Lacko, S., & Thornicroft, G. (2013). Mental Illness Stigma, Help Seeking, and Public Health Programs. *American Journal of Public Health*, *103*(5), 777–780. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3698814/

Hinshaw, S. P., & Stier, A. (2008). Stigma as related to mental disorders. *Annual Review of Clinical Psychology*, *4*(4), 367–393. https://doi.org/10.1146/annurev.clinpsy.4.022007.141245

Hizbul Khootimah Azzaakiyyah. (2023). The Impact of Social Media Use on Social Interaction in Contemporary Society. *Technology and Society Perspectives*, *1*(1), 1–9. https://doi.org/10.61100/tacit.v1i1.33

Hogeweg, L., & Neuleman, M. (2022). The hurtfulness of slurs, nouns and adjectives as group labels. *Linguistics in the Netherlands*, *39*, 72–87. https://doi.org/10.1075/avt.00062.hog

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Jacobs, L. (2008, May 21). *Judaism and Mental Illness*. My Jewish Learning; Oxford University Press. https://www.myjewishlearning.com/article/judaism-and-mental-illness/

Kallivayalil, R., & Enara, A. (2016). Stigma toward psychiatric disorders - National and International perspectives. *Indian Journal of Social Psychiatry*, *32*(2), 115. https://doi.org/10.4103/0971-9962.181096

Kemp, S. (2023). Digital 2023 October Global Statshot Report. In *DataReportal – Global Digital Insights*. Kepios. https://datareportal.com/reports/digital-2023-october-global-statshot

Kherwa, P., & Bansal, P. (2018). Topic Modeling: A Comprehensive Review. *ICST Transactions on Scalable Information Systems*, *0*(0), 159623. https://doi.org/10.4108/eai.13-7-2018.159623

Kielty, M. K. (2023, March 21). *John Taylor Recalls His Guns N' Roses and Sex Pistols Supergroup*. Ultimate Classic Rock. https://ultimateclassicrock.com/john-taylor-neurotic-outsiders/

Klára Látalová, Ján Praško, D. Kamarádová, Ocisková, M., A. Cinculova, A. Grambal, Radim Kubínek, B. Mainerova, Jarmila Smoldasova, Anežka Ticháčková, & Z. Sigmundova. (2014). Self-stigma and suicidality in patients with neurotic spectrum disorder - a cross sectional study. *PubMed*, *35*(6), 474–480.

Kristien Margi Suryaningrum. (2023). Comparison of the TF-IDF Method with the Count Vectorizer to Classify Hate Speech. *Engineering, Mathematics and Computer Science Journal (EMACS)*, *5*(2), 79–83. https://doi.org/10.21512/emacsjournal.v5i2.9978

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, *64*(4), 241–256. https://doi.org/10.1037/a0015309

Lee, K., Palsetia, D., Narayanan, R., Patwary, Md. M. A., Agrawal, A., & Choudhary, A. (2011). Twitter Trending Topic Classification. *2011 IEEE 11th International Conference on Data Mining Workshops*. https://doi.org/10.1109/icdmw.2011.171

Li, A., Jiao, D., Liu, X., & Zhu, T. (2020). A Comparison of the Psycholinguistic Styles of Schizophrenia-Related Stigma and Depression-Related Stigma on Social Media: Content Analysis. *Journal of Medical Internet Research*, *22*(4), e16470. https://doi.org/10.2196/16470

Lilienfeld, S. O., SauvignÃ©, K. C., Lynn, S. J., Cautin, R. L., Latzman, R. D., & Waldman, I. D. (2015). Fifty psychological and psychiatric terms to avoid: a list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01100

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M. S., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1907.11692

Loureiro, D., Barbieri, F., Neves, L., Espinosa Anke, L., & Camacho-collados, J. (2022, May 1). *TimeLMs: Diachronic Language Models from Twitter*. ACLWeb; Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-demo.25

McGarry, P. (1990). Misuse of psychiatric terms. *The British Journal of Psychiatry*, *156*(3), 444–444. https://doi.org/10.1192/bjp.156.3.444a

McGrory, A. (1980). Women and Mental Illness: A Sexist Trap? *Journal of Psychiatric Nursing and Mental Health Services*, *18*(9), 13–19.

McInnes, L., Healy, J., & Melville, J. (2020, September 17). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. ArXiv.org. https://doi.org/10.48550/arXiv.1802.03426

Moriceau, V., Benmara, F., & Boumadane, A. (2022). Automatic Detection of Stigmatizing Uses of Psychiatric Terms on Twitter. *13th Conference on Language Resources and Evaluation (LREC 2022), European Language Resources Association (ELRA)*, 237–243.

Nachar, N. (2008). The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 13–20. https://doi.org/10.20982/tqmp.04.1.p013

Naslund, J. A., & Deng, D. (2021). Addressing mental health stigma in low-income and middle-income countries: A new frontier for digital mental health. *Ethics, Medicine and Public Health*, *19*, 100719. https://doi.org/10.1016/j.jemep.2021.100719

Observatoire de la Société et de la Consommation (L'Obsoco). (2015). L'image de la schizophrénie à travers son traitement médiatique: Trois conclusions et une hypothèse. Etude lexicographique et sémantique. *Mandatée Par l'Association de Parents PromesseS, Membre Du Collectif National Schizophrénies*.

Padilla Cruz, M. (2019). Qualifying insults, offensive epithets, slurs and expressive expletives. *Journal of Language Aggression and Conflict*. https://doi.org/10.1075/jlac.00023.cru

Pavlova, A., & Berkers, P. (2020a). "Mental Health" as Defined by Twitter: Frames, Emotions, Stigma. *Health Communication*, *37*(5), 1–11. https://doi.org/10.1080/10410236.2020.1862396

Pavlova, A., & Berkers, P. (2020b). Mental health discourse and social media: which mechanisms of cultural power drive discourse on Twitter. *Social Science & Medicine*, *263*, 113250. https://doi.org/10.1016/j.socscimed.2020.113250

Pfeffer, J., Matter, D., Kokil Jaidka, Onur Varol, Afra Mashhadi, Lasser, J., Assenmacher, D., Wu, S., Yang, D., Brantner, C., Romero, D. M., Otterbacher, J., Schwemmer, C., Joseph, K., Garcia, D., & Morstatter, F. (2023). Just Another Day on Twitter: A Complete 24 Hours of Twitter Data. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2301.11429

Pinero, J. M. L. (1983). The concept of neurosis. In D. Berrios (Trans.), *Historical Origins of the Concept of Neurosis* (pp. 1–24). Cambridge University Press.

Reimers, N., & Iryna Gurevych. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. https://doi.org/10.48550/arxiv.1908.10084

Rentfrow, P. J., Jokela, M., & Lamb, M. E. (2015). Regional Personality Differences in Great Britain. *PLOS ONE*, *10*(3), e0122245. https://doi.org/10.1371/journal.pone.0122245

Rizaee, M. (2022). *Investigating the Potential of Topic Modelling On Social Media to Support the Identification of Refugee Needs* [MSc Thesis]. https://publications.scss.tcd.ie/theses/diss/2022/TCD-SCSS-DISSERTATION-2022-117.pdf

Rose, D., Thornicroft, G., Pinfold, V., & Kassam, A. (2007). 250 labels used to stigmatise people with mental illness. *BMC Health Services Research*, *7*(1). https://doi.org/10.1186/1472-6963-7-97

Salehabadi, N., Groggel, A., Singhal, M., Roy, S. S., & Nilizadeh, S. (2020). *User Engagement and the Toxicity of Tweets*. 18. https://doi.org/10.48550/arXiv.2211.03856

Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives. *Advances in Science, Technology and Engineering Systems Journal*, *2*(1), 127–133. https://doi.org/10.25046/aj020115

Sample, I. (2015, March 25). Smiling Scots, worried Welsh and lazy Londoners: survey maps regional personality types. *The Guardian*. https://www.theguardian.com/science/2015/mar/25/survey-maps-regional-personality-types

Schmid, U. K., Kümpel, A. S., & Rieger, D. (2022). How social media users perceive different forms of online hate speech: A qualitative multi-method study. *New Media & Society*, 146144482210911. https://doi.org/10.1177/14614448221091185

Silva, Juliana Tarossi Pollettini, & Antonio Pazin Filho. (2023). Unsupervised natural language processing in the identification of patients with suspected COVID-19 infection. *Cadernos de Saúde Pública*, *39*(11). https://doi.org/10.1590/0102-311xen243722

Silverman, M. H., Wilson, S., Ramsay, I. S., Hunt, R. H., Thomas, K. M., Krueger, R. F., & Iacono, W. G. (2019). Trait neuroticism and emotion neurocircuitry: Functional magnetic resonance imaging evidence for a failure in emotion regulation. *Development and Psychopathology*, *31*(3), 1085–1099. https://doi.org/10.1017/s0954579419000610

Sims, A. (1983). *What is Neurosis?* 1–24. https://doi.org/10.1007/978-1-349-17113-2_1

Singhal, R., & Rana, R. (2015). Chi-square Test and Its Application in Hypothesis Testing. *Journal of the Practice of Cardiovascular Sciences*, *1*(1), 69. https://doi.org/10.4103/2395-5414.157577

Stableford, D. (2020, August 13). Trump rants about his women troubles: Kamala, Mika, Nancy and AOC. *Yahoo!News*. https://ca.sports.yahoo.com/news/trump-sexist-attacks-mad-nasty-kamala-harris-aoc-pelosi-mika-crazy-ditzy-wife-165213724.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmN

vbS8&guce_referrer_sig=AQAAAKSvydqkojN_2VTycpnRaZ8mAM8c9fUxjFHvEBzjvQ3qa3rjaxZxSEVjNABMi63JfU60WOjXG8unOK3rErppYBEUTlUFNTds6r4t3lR33DCs8V-PAX9Q73f1tUFhWgcIXG6W9eaSQgREjjrZ0yJRufOgpr0eNDA6LY06bkv_N4Wz

Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, *7*(3), 321–326. https://doi.org/10.1089/1094931041291295

Szcześniak, D., Kobyłko, A., Lenart, M., Karczewski, M., Cyran, A., Musiał, P., & Rymaszewska, J. (2021). Personality Factors Crucial in Internalized Stigma Understanding in Psychiatry. *Healthcare*, *9*(4), 456. https://doi.org/10.3390/healthcare9040456

Tasca, C., Rapetti, M., Carta, M. G., & Fadda, B. (2012). Women and Hysteria in the History of Mental Health. *Clinical Practice & Epidemiology in Mental Health*, *8*(1), 110–119. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3480686/

Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The Dynamics of Political Incivility on Twitter. *SAGE Open*, *10*(2), 215824402091944. https://doi.org/10.1177/2158244020919447

Van Corbach, H. (2020). *AM Cluster Manual AM Cluster Manual*.

Walsh, D. A. B., & Foster, J. L. H. (2021). A Call to Action. A Critical Review of Mental Health Related Anti-stigma Campaigns. *Frontiers in Public Health*, *8*(569539). https://doi.org/10.3389/fpubh.2020.569539

Walten, A. G. (2015, August 27). Does Neuroticism Breed Creativity? Study Says "Yes." *Forbes*. https://www.forbes.com/sites/alicegwalton/2015/08/27/does-neuroticism-breed-creativity/?sh=3daec5ed618f

Weed, N. C., & Kwon, S. (2019). Neuroticism | psychology. In *Encyclopædia Britannica*. https://www.britannica.com/science/neuroticism

Widiger, T. A., & Oltmanns, J. R. (2017). Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry*, *16*(2), 144–145. https://doi.org/10.1002/wps.20411

Winberg, O. (2017). Insult Politics: Donald Trump, Right-Wing Populism, and Incendiary Language. *European Journal of American Studies*, *12*(2). https://doi.org/10.4000/ejas.12132

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political Polarization on the Digital Sphere: a Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication*, *38*(1-2), 1–42.

Zucker, R. (2016). *The Effects of Social Media on Music Sharing*. https://doi.org/10.33015/dominican.edu/2016.cms.st.01

Appendix A

Hyperlinks

Python code

Masterthesis.ipynb

R code

chi square test and linear regression

Topic Visualisations of Tweets containing Neurotic

Neurotic topic details (number, label, count, percentage and keywords)

Neurotic topics over time

Neurotic term rank

Neurotic intertopic distance map

Neurotic simple hierarchy topics

Neurotic hierarchy topics

Neurotic topic similiarity matrix

Neurotic topic keyword scores

Topic Visualisations of Tweets containing Neuroticism

Neuroticism topic details (number, label, count, percentage and keywords)

Neuroticism topics over time

Neuroticism intertopic distance map

Neuroticism term rank

Neuroticism simple hierarchy topics

Neuroticism hierarchy topics

Neuroticism topic similarity matrix

Neuroticism topic keyword scores

Appendix B

Table B1

Variance measures of mean likelihood of offensive content over the years for NT and NCMT

Measure	NT	NCMT
Min (Date)	0.25 (May 13, 2015)	0.16 (August 11, 2015)
Max (Date)	0.42 (July 10, 2021)	0.29 (July 15, 2020)
Mean	0.37	0.24
Standard Deviation	0.03	0.02

Appendix C

Table C1

Running Time of the Offensive Language Detection and Topic Modelling

Dataset	Offensive Language Detection	Topic Modelling
Neurotic	11 h 31 min	6 h 24 min
Neuroticism	55 min	33 min