Closed Dictionary, Language Based Personality Prediction

Investigating Personality Trait Expression in Twitter Interactions

Aleksandrs Baskakovs (202008101, AB), Luke Ring (202009983, LR)

Faculty of Arts, Aarhus University

Cognition and Communication (147201U007); Exam Code 147201E009

Bachelor of Cognitive Science

February 11, 2021

Abstract

This paper employs a natural language processing model for personality extraction from text to investigate the expression of the "neuroticism" personality trait in interactions between Twitter users and influencers of same and opposite political leanings.

The model utilized correlations between personality traits and LIWC dictionaries from Schwartz et al. (2013), which was derived from 700 million words from posts by Facebook users who took the Revised NEO Personality Inventory test. The model was validated by collecting publicly shared personality test results of various Twitter users (N = 52) and comparing them with model-predicted scores. A statistically significant correlation was found between the model's predictions and the users' reported scores for neuroticism.

Further employment of the model showed a statistically significant increase in predicted neuroticism scores for users' tweets (N = 17,885) directed at left-leaning as opposed to right-leaning influencers. Influencers (N = 12) were selected based on popularity, gender, and political leaning and were used to identify user accounts that mention them above a frequency threshold. Content-based sentiment analysis was performed against tweets that contained specific keywords to categorize users as politically left or right. These findings provide partial validation for natural language processing-based personality modelling and suggest further areas for an investigation into influencer and user language alignment as well as personality modelling in general.

Keywords: Twitter, social media, personality modelling, natural language processing

Closed Dictionary Language Based Personality Prediction

Investigating Twitter Users and Language Alignment

Due to the massive scale and complexity of online interactions, many potential avenues of research into human behaviour become available by analysing publicly posted texts. Finding correlations between individuals' interactions and their known behaviours has provided the ability to highlight potential risks for self-harm (Coppersmith et al., 2018), and has been used in attempts to predict election results (Metaxas et al., 2011) and even identify adverse side-effects from medication (Nikfarjam et al., 2015). The opinion of this paper's authors is that the importance of language analysis will continue to increase for understanding human behaviour.

Personality (AB, LR)

In the information age, the consumer, having dealt with the issue of content availability is facing a new problem - the paradox of choice (H. Hu & Krishen, 2019). With hundreds of thousands of varieties of products and services, which is the correct one? The effectiveness of a particular piece of content cannot be valued without first examining how well the content fits with the prior knowledge of the consumer (Amadieu et al., 2009). However, due to the uniqueness of individual human beings, evaluation of potential responsiveness of consumers to a particular message, be it a quiz question or a newsletter title, has been mostly done by targeting simple demographic facets. Understanding the personality of a particular individual could enable finer content customisation and could in turn be crucial for innovation in such fields as Education Technology.

One of the ways to gain deeper insight into personality is through attempts to quantify certain aspects of it. Costa and McCrae's Revised NEO Personality Inventory (NEO-PI-R) is used to measure personality for the five-factor model (Costa & McCrae, 2008), and is widely used in personality research[1]. The respective traits of the five-factor model are Openness; Conscientiousness; Extraversion; Agreeableness; and Neuroticism, also referred to as OCEAN. A recent literature review of genome wide association studies found - especially with neuroticism - several loci correlations to scores for the five-factor model's personality traits (Sanchez-Roige et al., 2018), indicating some empirical support for these traits being measurable and meaningful categorisations. Although the categories in the five-factor model are broad, they have use for identifying aspects of personality, and some associated adjectives provided for each of the five traits aid in outlining the definitions of the traits themselves (Table 1).

Table 1

Five-factor Personality Traits and Associated Adjectives

Trait	Adjectives
Openness (O)	Artistic, Curious, Imaginative, Insightful, Original, Wide Interests
Conscientiousness (C)	Efficient, Organized, Planful, Reliable, Responsible, Thorough
Extraversion (E)	Active, Assertive, Energetic, Enthusiastic, Outgoing, Talkative
Agreeableness (A)	Appreciative, Forgiving, Generous, Kind, Sympathetic
Neuroticism (N)	Anxious, Self-pitying, Tense, Touchy, Unstable, Worrying

Note. Table data are from McCrae & John (1992).

Extracting personality from text (AB, LR)

In order to gain access to the personality characteristics of an individual consumer, they generally would need to complete some form of personality questionnaire. Moreover, the consumer would need to share this information with researchers. In addition, the time required to facilitate obtaining these data, an individual's personality traits can change over time (Bleidorn et al., 2009;

[1] A Google scholar search (January 2, 2021) for "neo-pi-r" reports 47,700 results with 14,486 citations of Costa & McCrae's publication (2008). See: https://scholar.google.com/scholar?hl=en&q=%22neo-pi-r%22

Roberts et al., 2006), meaning a re-test would be required periodically to maintain accurate information on individual personality trait scores. In the context of gaining access to personality characteristics of a consumer for better tailoring of information products, the approach of collecting and re-collecting questionnaire results is clearly not the most efficient method, nor is it easily scalable or consumer friendly.

Research has indicated that the personality traits of an individual seem to be related to their behaviour (Kennair et al., 2020) and as such, their language as well. Therefore, it must be possible to infer something about individuals' personality and where they lie on the respective five-factor model traits from a body of text they have produced. As Tauszczik & Pennebaker put it:

"Language is the most common and reliable way for people to translate their internal thoughts and emotions into a form that others can understand. Words and language, then, are the very stuff of psychology and communication." (Tausczik & Pennebaker, 2010)

The endeavour of locating correlations between language use and these five personality traits is not a new one (Dewaele & Furnham, 1999; Scherer, 1979), however one of the first researchers to bridge the world of computational linguistics and personality psychology were Pennebaker & King, who analysed patterns in word use in to formulate specific, predetermined word categories and compile a dictionary of these words. They found consistent results across domains, showing that, for example, people who score high on extraversion use more social process talk, positive emotion words and inclusives; and fewer negations, tentative words, exclusives, causation words, negative emotion words, and articles (Groom & Pennebaker, 2002; J. W. Pennebaker & King, 1999).

LIWC (AB, LR)

One of the most established textual analysis methods out there is the Linguistic Inquiry and Word Count (LIWC, pronounced "Luke"). The LIWC2015 dictionary consists of almost 6,400 words, word stems and select emotions, with each filled into one or more sub-dictionaries by human judges.

Sub-dictionaries represent one of the 74 word categories through which LIWC compiles a text (J. W. Pennebaker et al., 2015). For example, *Linguistic Pronouns* which include categories like *1st person singular*, *2nd person, articles; Other grammar* include categories like *common verbs, numbers* and *Psychological Processes* include categories like *Positive emotion, social processes, drives* and many others. On the level of specific words, the word “cried” is part of five word categories: *sadness*, *negative emotion*, *overall affect*, *verb*, and *past tense verb*. Hence, if it is found in the target text, each of these five sub-dictionary scale scores will be incremented.

Leveraging Natural Language Processing (AB, LR)

The above-mentioned method of analysing bodies of text using human-designed vocabularies is called a closed-vocabulary or closed-dictionary approach and as of 2020, is the most common approach to text analysis, with LIWC being the most popular[2] dictionary within closed-vocabulary approaches. However, in the last two decades, thanks to the advent of big data and widespread availability of computation power, data-driven open-vocabulary approaches have been gaining popularity, especially within the field of Computer Science. Some of the more widely used open-vocabulary methods include Latent Semantic Analysis or LSA (Landauer & Dutnais, 1997), Latent Dirichlet Allocation or LDA (Blei et al., 2003) and word embeddings or Word2Vec (Mikolov et al., 2015).

In contrast to closed-vocabulary methods, where human-designed and psychological theory based dictionaries are used to analyse the text, open-vocabulary methods employ a data-driven, algorithmic approach to find patterns in the data. When closed and open vocabulary approaches have been compared in research, both approaches have benefits and drawbacks. Specifically, one example of a potential drawback of the closed-vocabulary method used in this paper is that while the LIWC dictionary is extremely comprehensive, it is still based on words in isolation, which may lack

[2] As of December 2020, the three main versions of LIWC (2001: (J. Pennebaker et al., 2001); 2007: (J. Pennebaker et al., 2007); 2015: (J. W. Pennebaker et al., 2015)) were cited over 8000 times.

contextual information that could be gained through word-phrases or other methods that are context-aware (Eichstaedt et al., 2020).

Due to the widespread adoption and increasing usage of social media, there is now a wealth of textual data generated by humans that is publicly available to researchers. This data can also be deemed as naturalistic, as social media sites like Facebook and Twitter encourage users to record their thoughts, emotions and behaviours in real-time. Moreover, this means that these data can be aggregated and analysed over time (Kosinski et al., 2015). Analysing of this content alongside use patterns and multi-party interactions increases the potential to highlight individual changes that may serve as markers for behaviours. This availability of user-generated content makes for a compelling case of combining both closed-vocabulary and open-vocabulary approaches on big, naturalistic datasets sourced from social media platforms.

In the context of personality prediction, much investigative work has been accomplished by Schwartz et al. (2013). Their study analysed Facebook posts of almost 75,000 users who had also completed NEO Personality Inventory Revised (NEO-PI-R) based questionnaires, making it the largest study of personality and language to date, and serves as the core research upon which this paper was based. In addition to establishing results correlating LIWC2007 dictionary categories and Five Factor Model personality traits, they utilized an open-vocabulary approach to identify words and phrases that highly and significantly correlate with the respective five-factor model personality traits of openness, conscientiousness, extraversion, agreeableness and neuroticism, as well as age and gender. This included novel discoveries such as people who ranked low on neuroticism (emotional stability) were more likely to mention social sports and life activities, like church and basketball, and people who rank low on extraversion, were likely to refer to modern Japanese culture, like anime and manga.

A meta-analysis of various studies - including the study mentioned above - of predicting five-factor model personality traits from text concludes that personality traits can be inferred from social media data with remarkable accuracy (Azucar et al., 2018). The authors also state that this efficient, and, as concluded, effective approach opens a whole new realm of possibilities within various fields such as human-computer interaction by optimizing interface designs (Farnadi et al., 2016); increase the efficacy of implementation of public health programs via customisation of communication materials (Chapman et al., 2014; Franks et al., 2009); and presenting information that is tailored to resonate with the user based on their personality (Golbeck et al., 2011; Nass & Lee, 2000). Is it possible to build a personality prediction model by simply implementing the personality-language correlation data from previous studies?

Personality expression in online interactions (AB, LR)

Is our personality static or dynamic? Does our personality differ when our interactions are directed to specific users as opposed to the general public? Do we differ in how we express neuroticism when speaking to a person of an opposite political view?

The concept of personality traits itself implies that there is an enduring consistency in persons behaviour across situations, also known as cross-situational consistency of personality (Alston, 1975). This view, however, was criticised frequently under what is known as the person-situation debate in personality psychology (Kenrick & Funder, 1988). Tett & Guterman (2000) propose an *interactionist principle of trait activation*, which emphasizes the opportunity of expressing a particular trait to a higher degree depending on the context and the environment, at the same time preserving the cross-situational consistency of personality.

There is some indication that five-factor trait scores vary depending on the context, and when comparing online and offline communication results from personality questionnaires, Blumer and Döring (2012) found that neuroticism scores, in the context of online communication, were significantly lower. This was one of the areas of interest for this paper, as it is possible that analysis of different types of interactions from both individuals and groups would display context-specific differences in personality traits, especially for interactions between same-opinion and contradictingopinion individuals and groups. Given that differences in language and biases exist when dealing with in-group and out-group individuals (Maass et al., 1989), any analyses based on these types of interactions must reflect those differences and have an effect on model outcomes.

Sentiment analysis (AB, LR)

When analysing the interaction between people of different political views, how does one classify a person as, for example, Conservative or Democrat? One of the natural language processing techniques used in this paper was sentiment analysis, which is a systematic way to extract subjective opinions about a specific topic by analysing language use or by finding occurrences of certain words associated with emotions. These topic-associated emotions can be categorised into specific emotions with associated scores, such as joy or anger, or as a binary classification of positive or negative sentiment. Sentiment analysis is widely used in consumer research to identify public opinions regarding products and services and has been the topic of large amounts of academic research (Feldman, 2013). This paper uses sentiment analysis to classify a Twitter user as either right-leaning or left-leaning, depending on the sentiment expressed in users' Tweets mentioning particular Democratic or Conservative keywords. This paper also acknowledges the difficulties in classifying political opinions while accounting for sarcasm and multiple topics or sentiments in a single tweet, particularly in regard to predicting outcomes (Ebrahimi et al., 2017), the manual review process that was used suggested a sufficient level of accuracy for this investigation.

Methods

Content collection from Twitter (AB, LR)

Some content was collected using the Twitter API[3], though limitations in gathering search results and tweets from users with high volumes of content made that approach ineffective for data analysis. The majority of content collected from Twitter was gathered using the Open-Source

[3] https://developer.twitter.com/en/products/twitter-api

Intelligence tool TWINT (Cody Zacharias & Francesco Poldi, 2020) which allowed for a more comprehensive collection of the tweets required for this paper, as well as being able to be integrated into the python scripts written for automating the organisation and collation of content and user meta-data.

Pilot investigation (AB, LR)

Initial pilot data were collected from followers of several well-known Twitter accounts with a minimum of 1,000 posted tweets and over 1,000,000 followers in order to increase the likelihood of gathering sufficient data to analyse. Tweets from the influencers as well as tweets mentioning the influencers were collected and stored in flat-file databases. These data were indexed by username for tweets by a single user, or by influencer username if tweets were targeting a specific influencer. A python script was written to calculate the number of unique users that mentioned the influencers, as well as the frequency of mentions per user and the tweet content. This provided a way of identifying users that interacted with influencers at a sufficient level to analyse the language of those specific interactions.

Manual review of tweets mentioning the influencers from the highest frequency users revealed several bot (automated) accounts, parody accounts or spam tweets (the same message repeated multiple number of times), these accounts were removed from the corpus wherever they were identified. Subsequently the tweets per user mentioning a selected influencer were collated against the LIWC2015 dictionary to create document feature matrices using the *quanteda* text analysis package (Benoit et al., 2018). Visualisations of the data were performed with z-standardised relative frequencies of LIWC categorised words, these visualisations were used for exploration of potential options for grouping influencers and users based on traits such as influencer gender, follower counts, and the number and type of accounts users followed. As no reliable automated ways to do these categorisations were feasible within the time constraints, the investigation focussed on the influencer and user political leaning.

After successfully categorising language from the collected tweets, beta values from Schwartz et al.'s (2013) correlation table were used to calculate predicted scores for all traits on the five-factor model as well as the additional predictions for age and gender. Given that age data was not readily available for Twitter users, this was excluded from further modelling. Manual review of the model's predicted gender revealed extremely inconsistent and inaccurate predictions and whilst alternative (i.e., non-language-based analysis) methods have been proposed for classification of gender such as using public profile information (Alowibdi et al., 2013) this was considered out of scope for this paper.

Personality Prediction Model (AB, LR)

The methodological basis for this open investigation was the published work of multiple collaborating researchers modelling language use and personality scores (Schwartz et al., 2013). This paper utilized their findings to implement a language-based personality prediction model, or, specifically, five different models - one for each of the traits - Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness.

Reproducing Closed Vocabulary Personality Prediction

As the data required to build an exact copy of Schwartz et al.'s LIWC-based model were not published due to ownership of the data as well as privacy concerns, the model for this research was based purely on the LIWC-personality correlation table provided in the paper (2013, fig. 2)[4] which detailed each of the standardized multivariate regression coefficient values for individual LIWC categories and the five-factor personality dimensions. One difference of note was that Schwartz et al. used the LIWC2007 dictionary for classification whereas the model in this paper used the updated LIWC2015, as the changes were well documented it was possible to map the new categories based

---

[4] Direct link to the table: https://doi.org/10.1371/journal.pone.0073791.g002

on the LIWC2015 manual (James W. Pennebaker et al., 2015) and the mapping performed for this research is described in Table 2.

Table 2

Categories Mapped from LIWC2007 to LIWC2015

LIWC2007	LIWC2015
inhibition	—
—	adj, affiliation, compare, differ, drives, informal, interrog, netspeak, power, reward, risk
inclusive	conjunction
exclusive	differentiation
humans	female
future	focusfuture
past	focuspast
present	focuspresent
humans	male

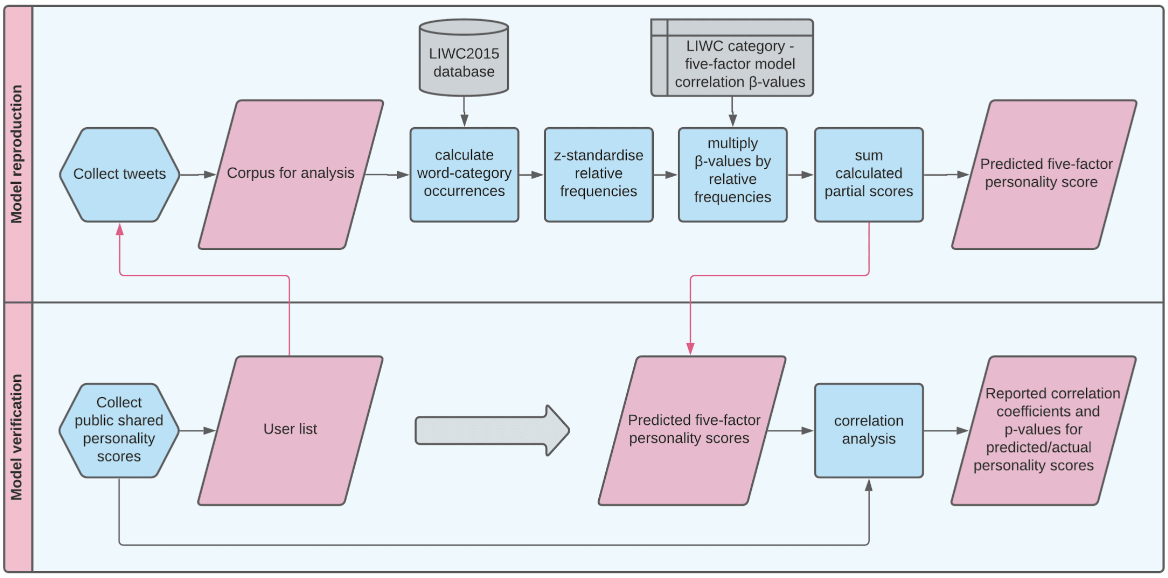
Note. Categories with a corresponding bar (—) were excluded from the model.

The reported beta values for each personality trait-LIWC category correlation were used as linear weights for overall personality predictions. Specifically, a corpus of text was formed from the collected tweets. The quanteda text analysis package (Benoit et al., 2018) was used to determine LIWC categorised word frequencies within that corpus. From this, z-standardised relative frequencies were calculated across all LIWC categories which were multiplied by their corresponding beta values to give a partial score. The sum of each category's partial score within a given dimension was used to produce a final score for the corresponding personality dimension.

A visual representation of the model's reproduction and verification process is shown in Figure 1. In addition, the R and python source code for the analysis and visualisation scripts are available on GitHub[5].

Figure 1

Closed-dictionary personality prediction model process



Evaluating the Model

The correlation table shows that the personality prediction model was derived from a very large dataset of at least 700 million words. However, as quantity alone is insufficient for the scores of the model to be trusted, the model was validated prior to performing analysis. To achieve this, the accuracy of the personality prediction model was benchmarked against the results of people who had taken the test and shared their results publicly via Twitter. The tweets of these users were analysed, and the model was used to make a prediction for each of the five personality traits. These predictions were compared to the actual five-factor model personality results that were shared by the users.

[5] https://github.com/zeyus/Twitter-NLP-politics-personality

After getting the predicted personality scores for each of the participants in the test sample, the normality of the scores was first assessed and transformed where required. Following that a fitting correlation analysis was applied to see whether changes in real scores are significantly mirrored in the predicted scores from our model.

Data Collection

Utilizing Twitter's Advanced Search functionality[6], 52 English-speaking Twitter users were found who have publicly made their five-factor personality test scores available through either https://your-persona.com/ or https://bigfive-test.com/. The posted personality scores were manually copied into a spreadsheet and standardized to be on a continuous scale from 1 to 100. Gender distribution was as follows: 23 male users, 27 female users and 2 non-binary users. The tweets of the users used for personality prediction were collected via Twitter v1.1 API[7]. Following that, up to 3100 of the users' latest tweets were fetched and the personality prediction model was applied for each of the five traits.

Utilizing the Model for Five-Factor Scores (AB, LR)

Data Collection

Influencers were selected from well-known public United States left- and right-leaning political Twitter users, who have publicly available content on YouTube or other media to increase the likelihood of other users seeing their content and interacting with them. Influencers were excluded who had less than 100,000 followers on Twitter, less than 1,000 tweets, or who were active politicians. A total of 12 accounts were included, with equal distribution of female left (n = 3), male left (n = 3), female right (n = 3) and male right (n = 3) accounts (Table 3).

[6] https://twitter.com/search-advanced

[7] https://developer.twitter.com/en/products/twitter-api

Tweets from the influencers and tweets mentioning the influencers - excluding links, replies and retweets[8] - were collected, stored in flat file databases and indexed by the username of the influencer mentioned. A python script was written to aggregate the content and create a list of the top 60 (or all, whichever was less) users by influencer mention frequency, with a lower limit of 10 mentions. This produced a list of 475 users matching the selection criteria that had mentioned one or more of the selected 12 influencers. The tweets from these users to the influencers (*N* = 61,197) were used in the analysis.

Table 3

Selected Twitter Influencers and Relevant Demographic Information.

Username	Followers	Tweets[9]	Political Leaning	Gender
BenShapiro	3,508,423	>162,400	Right	Male
ContraPoints	378,664	5,041	Left	Female
glennbeck	1,411,353	>19,100	Right	Male
hbomberguy	292,575	>25,800	Left	Male
IngrahamAngle	3,983,037	>46400	Right	Female
KellyannePolls	3,629,393	9,103	Right	Female
MsBlaireWhite	315,421	2,154	Right	Female
PhilosophyTube	187,181	>13,700	Left	Male
RubinReport	968,999	>85,300	Right	Male
shoe0nhead	404,959	>56,900	Left	Female
TheLindsayEllis	298,910	>14,200	Left	Female
VaushV	133,229	>14,200	Left	Male

Note. Follower tweet counts and demographic information were recorded January 3, 2021.

[8] Search pattern: (@<username>) -filter:links -filter:replies

[9] Exact number of tweets are not given for counts above 10,000 and are displayed as rounded values.

Determining Follower Political Alignment

In order to compare neuroticism scores between Twitter users based on *user to mentioned influencer* political leaning pairs, users were categorised into left or right political leanings using sentiment analysis of tweets containing specific keywords.

The United States 2020 presidential elections generated a huge amount of social media activity surrounding it with twitter alone seeing 965,620,919 related tweets (Chen, 2020; Chen et al., 2020). As such, the search keywords “JoeBiden” and “democrat” were selected for politically left terms and “realDonaldTrump” and “republican” were selected for politically right terms for analysing sentiment of users’ tweets on the political spectrum. Tweets since January 1, 2020 were gathered from search results for the four keywords for each of the 475 matching users[10], resulting in 197,512 unique tweets from 392 users after removing retweets, URLs and stop words from tweet content. Sentiment scores per user were assigned based on the “Bing” dataset (M. Hu & Liu, 2004) which gives words a related sentiment of either *positive* or *negative*. Relative frequencies of positive and negative sentiments for both right and left targets were calculated with ranges from -1 to +1. A political leaning score was calculated by subtracting users’ left score from their right score, giving values from -2 (most positive sentiment to left keywords and most negative sentiment to right keywords) to +2 (most positive sentiment to right keywords and most negative sentiment to left keywords). Users with less than 30 samples, or with absolute political leaning scores < 0.1 were removed, leaving 17,885 tweets from 197 users.

[10] Twitter search term: <keyword> (from:<username>)

Results

Personality Prediction Model (AB, LR)

The model was assessed on each of the five personality dimensions in the five-factor model for each of the 52 users with publicly available personality scores.

Testing assumption of normality

The data for both predicted and reported personality trait scores were assessed for normality using the Shapiro-Wilk test. All of the scores (predicted and actual) for all the personality traits met the assumption of normality (*skew2SE* < 1, *kurt2SE* < 1, $p > .05$) apart from the predicted score on the *openness* trait.

Correlation Analysis for Normally Distributed Traits

Correlation analysis was performed on the normally distributed data using Pearson's correlation test. Table 4 describes the results of correlation tests run on each of the traits.

Table 4

Pearson's Correlation Test Results for Four of the Five Traits.

Trait	*p*	Pearson's r	Result
Conscientiousness	.13	0.213	Non-significant
Extraversion	.12	0.217	Non-significant
Agreeableness	.79	0.037	Non-significant
Neuroticism	.04*	0.275	Significant

Note. * denotes significant results $p < .05$.

Pearson's correlation analysis was used to test model performance and found a statistically significant correlation between the model's predicted neuroticism scores and Twitter users' reported neuroticism scores, $r = 0.275$, $p = .049$, approaching the result of the original model[11], whilst utilizing fewer predictors and not being able to adjust for age and gender.

Correlation Analysis for Non-Normally Distributed Traits

The data for both predicted and reported openness scores were assessed for normality using the Shapiro-Wilk test, the predicted scores differed significantly from normal distribution (*skew2SE* > 1, *kurt2SE* > 1, $p < .05$), whilst the data for actual scores did not differ significantly from a normal distribution (*skew2SE* < 1, *kurt2SE* < 1, $p > .05$). After attempting both log and sqrt data transformations on the predicted scores and finding no improvement in normality, correlation analysis was performed using Kendall's tau non-parametric compatible test, selected over Spearman's due to the limited sample size, making Kendall's tau a better fit for the data. Test results indicated no significant correlation between the model's predicted openness scores and the reported scores from Twitter users, *tau* = 0.108, $p = .262$.

Investigating Emotional Stability in Political Leaning Based Interactions (AB, LR)

Having achieved significantly valid model performance on the neuroticism personality trait, the subsequent investigation focussed the expression of this personality trait in Twitter interactions between people of opposing and similar political leanings.

Users' Political Leaning

To verify the political leanings of the 197 users classified through the sentiment analysis process, 10 random users, plus the top and bottom 10 users ordered by their absolute political leaning scores were manually verified as accurate in the categorization of their tweets as left- or right-leaning.

---

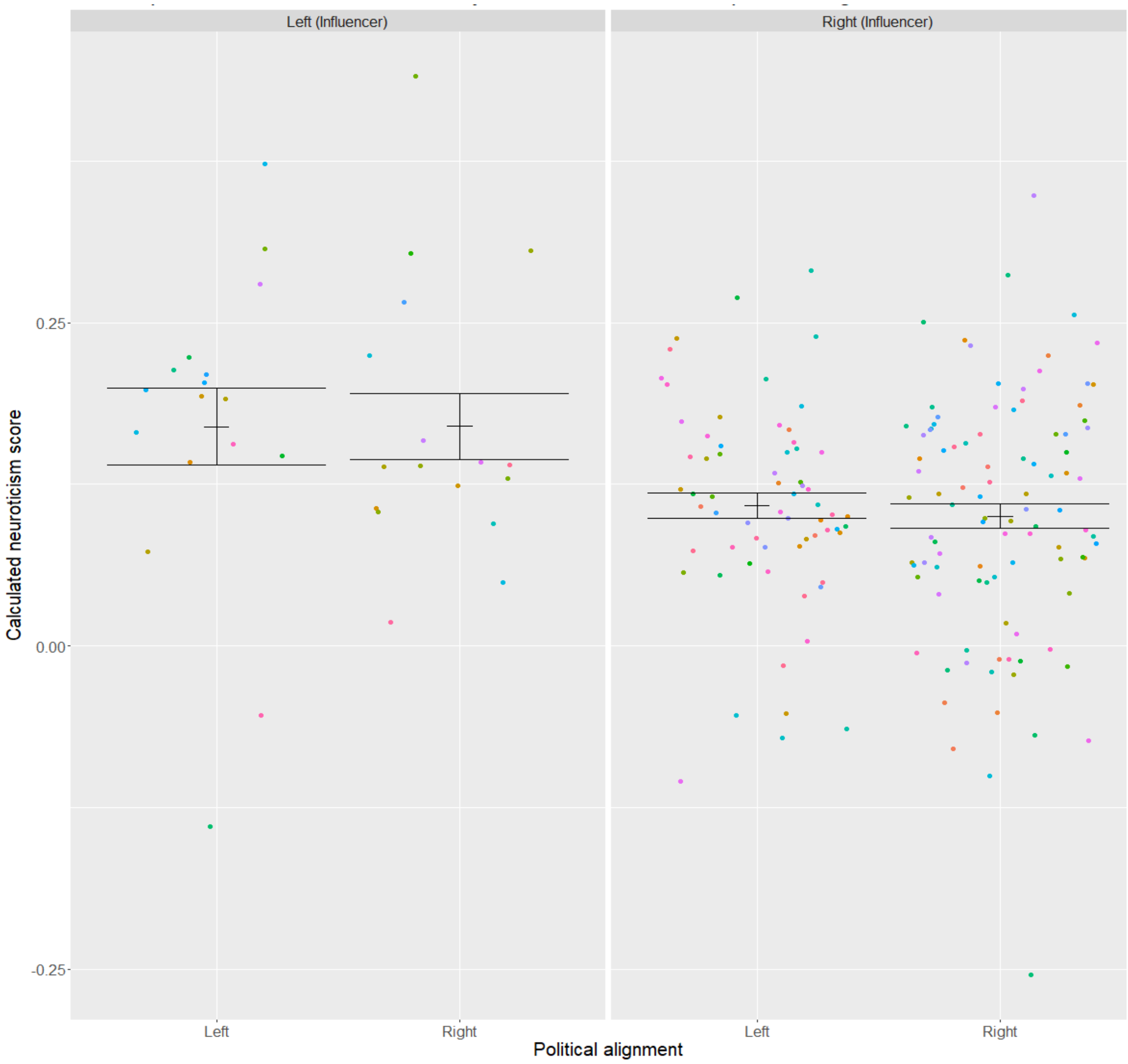
[11] The original study (Schwartz et al., 2013) achieved a Pearsons correlation of r > 0.3 across all five traits. This correlation is considered as "high" in the context of psychology research (Meyer et al., 2001).

Comparing Neuroticism Scores by Influencer and User Political Leaning

Based on the analysis of the predictions of the model for the five personality traits, the investigation focussed on Neuroticism as it achieved statistical significance. The other four traits were excluded from further analysis.

Figure 2

Scatter plot of user and target influencer neuroticism



Note. Calculated neuroticism scores for tweets containing mentions of the selected influencers by influencer and user political alignment, bars show means with standard error. *N* = 197. Left - Left (influencer): *n* = 17, *M* = 0.170, *SE* = 0.030. Right - Left (Influencer): *n* = 17, *M* = 0.170, *SE* = 0.026. Left - Right (influencer): *n* = 65, *M* = 0.109, *SE* = 0.010. Right - Right (influencer): *n* = 98, *M* = 0.100, *SE* = 0.010.

The determined influencer and user political alignments were used to create influencer-user alignment pairs to analyse the language and generate predicted neuroticism scores between these groups. The four pair groups for user-influencer political alignment were: 1) left-left, 2) right-left, 3) left-right, 4) right-right. All of the individual users' (N = 173) tweets (N = 17,885) to the influencer were analysed as a corpus per user, and their scores were plotted (Figure 2). The data were visually inspected for normality and equality of variance and as they satisfied the assumptions, correlation was investigated using a one-way analysis of variance was performed using the formula '*neuroticism_score ~ user_pol_alignment + influencer_pol_alignment*', and the results (see Table 5) showed a statistically significant correlation between the target influencers' political alignment and the calculated neuroticism score of the users' language from the tweets. No significant effect was seen from the users' political alignments on the calculated neuroticism score.

Table 5

One-Way Analysis of Variance of Neuroticism Score Predicted by Influencer and User Political alignment

Effect	$\hat{\eta}_G^2$	*F(1, 194)*	*MSE*	*p*
User political alignment	.003	0.64	0.01	.424
Influencer political alignment	.066	13.79	0.01	< .001

Discussion

Predicting Personality (AB, LR)

In this exploratory study we attempted to create a personality prediction model based on an existing study that used Facebook posts of 75,000 users, who have also taken a five-factor model personality test, to locate patterns between word usage and personality traits (Schwartz et al., 2013). Our model was built using only the correlation betas between LIWC dictionaries and five personality traits and was tested on tweets published by users who have made their five-factor model scores public via sharing them on Twitter.

Test Sample

We were able to find 52 Twitter users who have posted their five-factor model results, were predominantly English-speaking and had a sufficient number of tweets necessary for our analysis. By doing a correlation analysis between predicted and actual scores, the only trait which predicted scores that significantly correlated with actual scores was neuroticism. Conscientiousness and extraversion were approaching significance, and agreeableness and openness were rather insignificant. The main advantage of this approach is that the users voluntarily shared their personality test results, and therefore the research did not require organizing voluntary participant involvement.

An increase in test sample size would further clarify the validity of the model and could push the two traits closer towards significant results, opening an opportunity of doing further research with them. A potential problem with using the public test results is that no demographic information was collected for those users, meaning that the beta values used in this model could not be adjusted for age and gender, which may have been a reason for the overall lower correlation scores and lack of statistical significance for the personality traits apart from neuroticism.

Facebook & Twitter differences

The original study (Schwartz et al., 2013) is based on Facebook post data, which slightly differs from the Twitter tweet data, with the most notable difference being the Twitter character cap of 280[12] total characters per tweet. Thus, we do not discard the possibility of domain adaptation issues with the model.

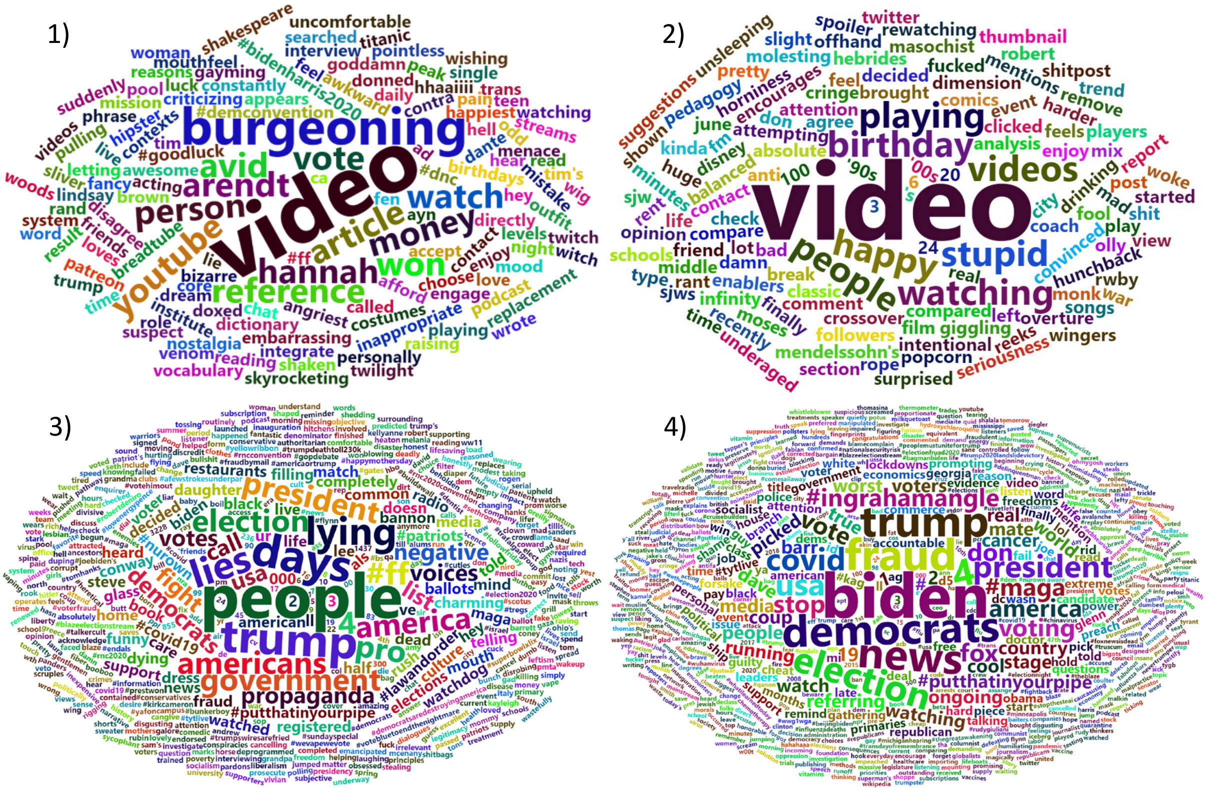
[12] 140 characters prior to November 2017.

Neuroticism and political alignment (AB, LR)

As this was an open-ended investigation without an initial research hypothesis, the gathered data were investigated for many potential correlations and areas of interest for further research and discovery. We had a slight expectation that alignments in both language use and personality would arise among users grouped by similar demographics. The data shows that Twitters users' predicted neuroticism score is significantly higher when the target influencer for the communication is politically left, when compared with communication to a target influencer who is politically right. We speculate that there may be several reasons for this difference including that it may be related to the underlying model's classification, as well as the words that appeared most frequently in the content analysed (Figure 3). These differences in the language used in the four groups would likely change with a larger sample size. We also note that this research was based on US influencers and English-speaking users and did not include cross-cultural samples.

Figure 3

Influencer-Follower Paired Word Clouds



Note. 1) Left user, left influencer, 2) right user, left influencer, 3) left user, right influencer, 4) right user, right influencer. Wordclouds generated with the *wordcloud2* R package (Lang, 2021).

Using sentiment analysis provided a rapid way to classify the political leaning of Twitter users based on the content of their tweets containing specific politically related keywords. While manual verification of the results was considered acceptable for the purpose of this investigation, if this research were to be reproduced with larger datasets or more inclusive keywords, a machine-learning model to verify classification accuracy in a quantifiable manner would be a more methodological approach. One limitation of the classification is the use of only two keywords per category, and while there were still substantial data for those alone, including more keywords and variations of them such as “Donald Trump”, “Trump”, “democrats” and “democratic party” may have given more rich data. Another consideration when using a keyword-based approach would be that both parties may be mentioned in a single tweet, and refining search terms to exclude terms outside of the category may result in a more accurate classification. Finally, the Bing sentiment database is limited to single words, and does not account for sarcasm, although that may have been mitigated slightly by excluding links, replies and retweets, as some Twitter users will be aware that an isolated tweet that is not part of, or a response to a larger conversation could be taken at face value.

Considerations for Further Research

From the knowledge acquired through this investigation, a list of reflections on the potential areas for further research are detailed below.

The investigation focused on exploring neuroticism in interactions from a group of followers to a particular influencer. While this aims to highlight the general behavioural stance of members of a group with a particular political leaning to a person of an opposing or similar political leaning, it would be interesting to explore differences in neuroticism expression on an individual level. This would require locating individual users that frequently interact with both users of opposing and similar views, as well as frequently posting on Twitter itself, and to estimate how highly they score on neuroticism when posting on their own, when interacting with a user of a similar view and a user of an opposing view.

Using natural language processing models to assess personality traits as in this paper allows for interesting segmentation not possible when using the traditional questionnaire-based methodology. While it is possible to create custom, domain-specific versions of the questionnaires as was done with asking participants specifically about their online interactions (Blumer & Döring, 2012), this is labour intensive and relies on participants accurately describing differences in each situation. If predictive models can achieve sufficient levels of accuracy, this can be done without requiring subjective self-reported analyses. The potential for this is to investigate how language, behaviours and predicted personality traits change between many different classes of interaction thus providing context-specific data and a way to compare them.

With personality traits likely to change over time (Bleidorn et al., 2009; Roberts et al., 2006), an exploration of temporal dynamics of personality traits is possible. Utilizing Twitter which provides open access to the full history of an individual's tweets this investigation would also be highly naturalistic. Event-based dynamics of personality traits (Tett & Guterman, 2000) are also possible to investigate using the same approach. This would, however, require manual mapping of events onto the timeline of gathered tweets or finding a published dataset of associated tweet IDs.

Another potential area of interest would be to apply this kind of personality modelling to the language groups themselves, rather than groups of individuals. This is a possible way to classify behaviours that are part of larger group dynamics. For consumer research, individual data is extremely valuable, but for other areas of research, such as intelligence, gathering metrics from group-based data would allow for correlating known behaviours of extremist groups with collective language used by the group. This may provide an avenue for flagging groups that may exhibit potentially dangerous behaviours.

A dedicated study to improve the performance of the personality prediction model utilizing additional predictors, as well as adjusting for age and gender may be a way to fine-tune the model. Achieving good personality prediction model performance on all of the five traits (openness, conscientiousness, extraversion, agreeableness and neuroticism) would open additional avenues for possible studies, with the ability of mixing and matching personality traits depending on the goal of the investigation.

Conclusion

This investigation showed significant correlation between predicted and reported neuroticism scores on the five-factor model with a closed-dictionary based model using predefined weights. Additionally, significantly higher levels of neuroticism were found in twitter users' interactions with politically left leaning influencers when compared to interactions with politically right leaning influencers.

While we achieved success in creating and utilising the model, because of the limitations noted in the discussion section, we are hesitant to conclude or infer any strong conclusions about the political groups that were analysed. We would advise extreme caution when drawing conclusions based on predictive models such as the one in this paper, without first addressing the stated limitations and triangulating the results with other research methods.

Using natural language processing to identify personality traits from publicly available content alone shows potential as a general tool for use in naturalistic behavioural observations as well as a potentially valuable tool for understanding how self-reported internal states and behaviours relate to language use. Natural language processing and big data analysis are fields that are continuing to expand with new research and - especially when combined with machine learning - are powerful methods gaining greater understanding into human cognition.

References

Alowibdi, J. S., Buy, U. A., & Yu, P. (2013). Empirical Evaluation of Profile Characteristics for Gender Classification on Twitter. *2013 12th International Conference on Machine Learning and Applications*, 365–369. https://doi.org/10.1109/ICMLA.2013.74

Alston, W. P. (1975). Traits, Consistency and Conceptual Alternatives for Personality Theory. *Journal for the Theory of Social Behaviour*, *5*(1), 17–48. https://doi.org/10.1111/j.1468-5914.1975.tb00341.x

Amadieu, Franck, Tamara van Gog, Fred Paas, André Tricot, and Claudette Mariné. "Effects of Prior Knowledge and Concept-Map Structure on Disorientation, Cognitive Load, and Learning." *Learning and Instruction* 19, no. 5 (October 2009): 376–86. https://doi.org/10.1016/j.learninstruc.2009.02.005

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150–159. https://doi.org/10.1016/j.paid.2017.12.018

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bleidorn, W., Kandler, C., Riemann, R., Angleitner, A., & Spinath, F. M. (2009). Patterns and sources of adult personality development: Growth curve analyses of the NEO PI-R scales in a longitudinal twin study. *Journal of Personality and Social Psychology*, *97*(1), 142–155. https://doi.org/10.1037/a0015434

Blumer, T., & Döring, N. (2012). Are we the same online? The expression of the five factor personality traits on the computer and the Internet. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *6*(3). https://doi.org/10.5817/CP2012-3-5

Chapman, B. P., Hampson, S., & Clarkin, J. (2014). Personality-informed interventions for healthy aging: Conclusions from a National Institute on Aging work group. *Developmental Psychology*, *50*(5), 1426–1441. https://doi.org/10.1037/a0034135

Chen, E. (2020). *2020 US Presidential Election Tweet IDs* (936fbf40207ebf113121678a95218dd4f4e5c2d7) [Python]. https://github.com/echen102/us-pres-elections-2020/tree/936fbf40207ebf113121678a95218dd4f4e5c2d7

Chen, E., Deb, A., & Ferrara, E. (2020). #Election2020: The First Public Twitter Dataset on the 2020 US Presidential Election. *ArXiv:2010.00600 [Cs]*. http://arxiv.org/abs/2010.00600

Cody Zacharias & Francesco Poldi. (2020). *TWINT - Twitter Intelligence Tool* [Python]. TWINT Project. https://github.com/twintproject/twint

Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural Language Processing of Social Media as Screening for Suicide Risk. *Biomedical Informatics Insights*, *10*, 117822261879286. https://doi.org/10.1177/1178222618792860

Costa, P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In *The SAGE Handbook of Personality Theory and Assessment: Volume 2—Personality Measurement and Testing* (pp. 179–198). SAGE Publications Ltd. https://doi.org/10.4135/9781849200479.n9

Dewaele, J.-M., & Furnham, A. (1999). Extraversion: The Unloved Variable in Applied Linguistic Research. *Language Learning*, *49*(3), 509–544. https://doi.org/10.1111/0023-8333.00098

Ebrahimi, M., Yazdavar, A. H., & Sheth, A. (2017). Challenges of Sentiment Analysis for Dynamic Events. *IEEE Intelligent Systems*, *32*(5), 70–75. https://doi.org/10.1109/MIS.2017.3711649

Eichstaedt, johannes C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C., Tobolsky, V., Smith, L. K., Buffone, A., Iwry, J., Seligman, M., & Ungar, L. H. (2020). *Closed and Open Vocabulary Approaches to Text Analysis: A Review, Quantitative Comparison, and Recommendations*. PsyArXiv. https://doi.org/10.31234/osf.io/t52c6

Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., Davalos, S., Moens, M.-F., & De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, *26*(2), 109–142. https://doi.org/10.1007/s11257-016-9171-0

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82–89. https://doi.org/10.1145/2436256.2436274

Franks, P., Chapman, B., Duberstein, P., & Jerant, A. (2009). Five factor model personality factors moderated the effects of an intervention to enhance chronic disease management self-efficacy. *British Journal of Health Psychology*, *14*(3), 473–487. https://doi.org/10.1348/135910708X360700

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 253–262. https://doi.org/10.1145/1979742.1979614

Groom, C. J., & Pennebaker, J. W. (2002). Words. *Journal of Research in Personality*, *36*(6), 615–621. https://doi.org/10.1016/S0092-6566(02)00512-3

Hu, H., & Krishen, A. S. (2019). When is enough, enough? Investigating product reviews and information overload from a consumer empowerment perspective. *Journal of Business Research*, *100*, 27–37. https://doi.org/10.1016/j.jbusres.2019.03.011

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, 168. https://doi.org/10.1145/1014052.1014073

James W. Pennebaker, Roger J. Booth, Ryan L. Boyd, & Martha E. Francis. (2015). *Linguistic Inquiry* *and Word Count: LIWC2015*. Pennebaker Conglomerates, Inc. http://liwc.wpengine.com/wp-content/uploads/2015/11/LIWC2015_OperatorManual.pdf

Kennair, L. E. O., Solem, S., Hagen, R., Havnen, A., Nysæter, T. E., & Hjemdal, O. (2020). Change in personality traits and facets (Revised NEO Personality Inventory) following metacognitive therapy or cognitive behaviour therapy for generalized anxiety disorder: Results from a randomized controlled trial. *Clinical Psychology & Psychotherapy*, cpp.2541. https://doi.org/10.1002/cpp.2541

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*(1), 23–34. https://doi.org/10.1037/0003-066X.43.1.23

Kosinski, M., Matz, S., Gosling, S., Popov, V., & Stillwell, D. (2015). Facebook as a Research Tool for the Social Sciences. *The American Psychologist*, *70*, 543–556. https://doi.org/10.1037/a0039210

Landauer, T. K., & Dutnais, S. T. (n.d.). *A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge*. 30.

Lang, Dawei. *Wordcloud2: Create Word Cloud by HtmlWidget*, 2021. https://github.com/lchiffon/wordcloud2.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*(6), 981–993. https://doi.org/10.1037/0022-3514.57.6.981

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Metaxas, P. T., Mustafaraj, E., & Gayo-Avello, D. (2011). How (Not) to Predict Elections. *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 165–171. https://doi.org/10.1109/PASSAT/SocialCom.2011.98

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*(2), 128–165. https://doi.org/10.1037/0003-066X.56.2.128

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. A. (2015). *Computing numeric representations of words in a high-dimensional space* (United States Patent No. US9037464B1). https://patents.google.com/patent/US9037464B1/en

Nass, C., & Lee, K. M. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 329–336. https://doi.org/10.1145/332040.332452

Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, *22*(3), 671–681. https://doi.org/10.1093/jamia/ocu041

Pennebaker, J., Chung, C., Ireland, M., Gonzales, A., & Booth, R. (2007). *The Development and Psychometric Properties of LIWC2007*.

Pennebaker, J., Francis, L., & Booth, R. (2001). Linguistic inquiry and word count: LIWC2001. *LIWC Operators Manual 2001*.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. https://doi.org/10.15781/T29G6Z

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

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Sanchez-Roige, S., Gray, J. C., MacKillop, J., Chen, C.-H., & Palmer, A. A. (2018). The genetics of human personality. *Genes, Brain and Behavior*, *17*(3), e12439. https://doi.org/10.1111/gbb.12439

Scherer, K. R. (1979). Nonlinguistic Vocal Indicators of Emotion and Psychopathology. In C. E. Izard (Ed.), *Emotions in Personality and Psychopathology* (pp. 493–529). Springer US. https://doi.org/10.1007/978-1-4613-2892-6_18

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Tett, R., & Guterman, H. (2000). Situation Trait Relevance, Trait Expression, and Cross-Situational Consistency: Testing a Principle of Trait Activation. *Journal of Research in Personality*, *34*, 397–423. https://doi.org/10.1006/jrpe.2000.2292