Machine Learning in Recruiting: Predicting Personality from CVs and Short Text Responses

Preprint · September 2023

DOI: 10.31234/osf.io/bc4v9

5 authors, including:

Eric Grunenberg
University of Münster

Heinrich Peters
Columbia University

Mitja D Back
University of Münster

This is the final author version (before journal's typesetting and copyediting) of the following article:

Grunenberg, E., Peters, H., Francis, M. J., Back, M. D., & Matz, S. C. (2024). Machine learning in recruiting: predicting personality from CVs and short text responses. *Frontiers in Social Psychology*, *1*, Article 1290295. https://doi.org/10.3389/frsps.2023.1290295

Machine Learning in Recruiting: Predicting Personality from CVs and Short Text Responses

[PREPRINT VERSION]

Grunenberg, E. [1,2+], Peters, H.[2], Francis, M. J. [3], Back, M.D.[1], & Matz, S. C.[2]

[1]University of Münster, Germany, [2]Columbia Business School, USA,

[3] Independent Researcher, Australia

Author Note

ORCID

Grunenberg, Eric: https://orcid.org/0000-0001-6296-468X
Peters, Heinrich: https://orcid.org/0000-0002-0571-6388
Back, Mitja D.: https://orcid.org/0000-0003-2186-1558
Matz, Sandra C.: https://orcid.org/0000-0002-0969-4403

This research was supported by a scholarship by the German Academic Scholarship Foundation and by the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research, Germany; project number: 01GK1801A). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript and the authors did not report any conflict of interest. We embrace the values of openness and transparency in science (Schönbrodt et al., 2015). We therefore follow the 21-word solution (Simmons et al., 2012), or refer to project documentations in the OSF.

We are grateful to Johannes Klinz and Helena Sydlik for their help proofreading and both the members of the BackLab in Münster and the CBS Lab at Columbia Business School for fruitful discussions about this research.

[+]Correspondence concerning this article should be addressed to E. Grunenberg (eric.grunenberg@uni-muenster.de), University of Münster, Institute for Psychology, Psychological Assessment and Personality Psychology, Fliednerstr. 21, 48149 Münster, Germany

CRediT:

Eric Grunenberg: Writing – original draft, Writing – review and editing, Conceptualization, Formal analysis, Visualization

Heinrich Peters: Supporting role in conducting Formal analysis, supporting role in conceptualization and supporting role in Writing: reviewing and editing

Matt J. Francis: Resources and Data Curation, supporting role in Conceptualization as well as Writing – reviewing and editing

Mitja Back: Writing – review and editing, Resources, Supervision

Sandra Matz: Writing – review and editing, Conceptualization, Resources, Supervision

Abstract

Assessing the psychological characteristics of job applicants - including their vocational interests or personality traits – has been a corner stone of hiring processes for decades. While traditional forms of such assessments require candidates to self-report their characteristics via questionnaire measures, recent research suggests that computers can predict people's psychological traits from the digital footprints they leave online (e.g., their Facebook profiles, Twitter posts or credit card spending). Although such models become increasingly available via third-party providers, the use of external data in the hiring process poses considerable ethical and legal challenges. In this paper, we examine the predictability of personality traits from models that are trained exclusively on data generated during the recruiting process. Specifically, we leverage information from CVs and free-text answers collected as part of a real-world, high-stakes recruiting process in combination with natural language processing to predict applicants' Big Five personality traits (N = 8,313 applicants). We show that the models provide consistent moderate predictive accuracy when comparing the machine learning-based predictions with the self-reported personality traits (average r = .25), outperforming recruiter judgments reported in prior literature. Although the models only capture a comparatively small part of the variance in self-reports, our findings suggest that they might still be relevant in practice by showing that automated predictions of personality are just as good (and sometimes better) at predicting key external criteria for job matching (i.e., vocational interests) as self-reported assessments.

Introduction

The future of work is changing rapidly: Automation, digitalization, and globalization are predicted to drive a shift in both our concept of work and the workforce itself (Acemoglu & Restrepo, 2019; Spencer, 2018). In the next ten to fifteen years, entire industries are predicted to vanish, others will emerge, and some will reinvent themselves. Job markets around the world will face large scale workforce transitions: More and more people will be searching for new jobs due to skill or organizational mismatches (Manyika et al., 2017). Many of these job seekers will be matched by third-party HR services such as online job boards and agencies (Lund et al., 2016). These platforms need scalable solutions to process applicants' data and match them to potential jobs.

Recent research suggests that predictive modeling using machine learning (ML) offers a promising approach to replacing time-consuming self-reports by predicting organizationally relevant traits about job seekers (e.g., cognitive ability, personality, or interests). Notably these assessments are not only relatively accurate but also cheap, unobtrusive, and scalable (Azucar et al., 2018; Matz & Netzer, 2017; Stachl et al., 2021). However, most of the existing research on ML-based assessments of personality has either focused on later stage hiring procedures, such as interviews (e.g., Hickman et al., 2022; Hickman et al., 2021) or relied on personal data obtained from social media (e.g., Hickman et al., 2019; Kosinski et al., 2013; Park et al., 2015; Youyou et al., 2015). While digital footprints extracted from social media data are often easy to access (e.g. Facebook Likes or status updates, Tweets, or LinkedIn profiles; Kosinski et al., 2015), they have several limitations when it comes to their application in the recruiting process. First, research on personality prediction from assessment center interviews indicates that personality prediction models trained on social media do not necessarily generalize well on data from personnel selection settings (Hickman et al., 2022; Hickman et al., 2021; Hickman et al., 2019). Second, using social media data in general might be problematic: Applicants might not have active social media profiles, might vary in the amount of data available or might be opposed to recruiters invading their privacy by accessing their social media profiles and seemingly unrelated personal data (Matz et al., 2020; Matz & Netzer, 2017; Yarbrough, 2018).

In this paper, we suggest an alternative approach that does not rely on external data, but instead leverages information that is generated early in the recruiting process itself. Specifically, we show that (i) ML-based models trained on readily available short text excerpts and CVs native to a third-party recruiting agency can predict Big Five personality traits and (ii) that these ML-based personality assessments in turn predict vocational interests with similar accuracy as self-reported personality. Taken together, our results suggest that companies and third-party services can benefit from considering predictive models trained on internal recruiting data to assess candidates' psychological characteristics, respecting candidates' privacy, and ultimately use these insights to proactively match them to suitable jobs.

The Future of Work and the Need for Proactive Job Matching

The combination of automation, digitalization and globalization has led to major changes in (i) the workplace, (ii) employment structures and (iii) the meaning of work individuals expect from their professions (Acemoglu & Restrepo, 2017, 2019; Anderson et al., 2017; Makridakis, 2017). Studies expect that by 2030 up to 375 million people might need to switch occupational sectors and re-train due to automation. At the same time, people's expectations regarding their work are changing (Manyika et al., 2017). Especially among younger adults there is a growing demand for meaningful work as well as a healthy work-life balance (Anderson et al., 2017; Lub et al., 2016; Twenge et al., 2010).

The shifting nature of work has two important implications for recruiting. First, job seekers have access to a wider pool of jobs and change jobs more frequently than before. Second, companies have access to a wider pool of applicants but might have to hire more frequently and assure that the applicants they hire derive meaning from the jobs they enter. Together, these shifts lead to a more dynamic, complex job market that requires efficient and scalable matching between job seekers (and their preferences) and companies (with their job requirements; i.e., person-job fit; Kern et al., 2019;

Kristof, 1996; Wilmot & Ones, 2021). Consequently, digital intermediaries such as online job boards, platforms, and agencies (both as proprietary part of organizations and as third-party services) will play an increasingly important role in the recruiting process as they streamline the matching between applicants and jobs (Allen et al., 2007; Cardoso et al., 2021; Manyika et al., 2017; Schaarschmidt et al., 2021).

The matching process typically considers two types of information: (1) applicants' formal characteristics, including education, prior job experience, current location, or expected salary, and (2) applicants' "softer" but less superficial psychological characteristics such as motivation, personality, organizational values, or vocational interests. Personality traits are among the most prominent psychological characteristics considered in the hiring process (Salgado & Fruyt, 2017). They capture relatively stable individual differences in the way that people think, feel and behave. The most prominent taxonomy of personality is the Five Factor Model (Goldberg, 1993) – or Big Five.

The Big Five have been associated with a wide variety of organizational outcomes, including job performance (Barrick & Mount, 1991; Barrick et al., 2001; Wilmot & Ones, 2021), leadership (Bono & Judge, 2004; Judge et al., 2002), team member effectiveness (Bell, 2007; Mount et al., 1998) and counterproductive work behavior (Berry et al., 2007), for a comprehensive overview, see Mount & Barick (2012). What is more, research on person-job- and person-organization-fit suggests that individuals are more motivated by and perform better in jobs that match their personality characteristics (Kristof-Brown & Guay, 2011; Ostroff & Zhan, 2012; Wilmot & Ones, 2021).

Automated Predictions of Personality

An important consideration for the use of softer psychological criteria in the recruiting process, is the ease with which such information can be obtained. Traditionally, psychological characteristics have been assessed with self-report questionnaires (Funder, 2009), which require applicants to indicate how much they agree with different statements (e.g., "I am the life of the party" as a measure of Extraversion). In the context of scalable recruiting intermediaries the use of self-reports is often hindered by the fact that applicants might not be motivated to complete lengthy assessments (Hausknecht et al., 2004; Ryan & Ployhart, 2000). This is particularly true for proactive matching approaches that provide proactive recommendations to candidates. Consequently, if psychological factors are to be integrated into scalable matching algorithms, there is a need to facilitate psychometric assessments that do not rely on direct input from the user. A promising approach to assessing organizationally relevant characteristics such as personality traits within digital recruitment and recommendation systems is the application of automated ML-based predictions from text. ML approaches promise a scalable and efficient alternative to self-reports or human judgments (e.g., Goretzko & Israel, 2022; Hickman et al., 2022; Sajjadiani et al., 2022; Sajjadiani et al., 2019). Text is a nearly universal part of application platforms, which typically collect text in the form of CVs, cover letters, self-descriptions, or text data on online profiles. Compared to questionnaire responses, free text allows applicants to express themselves in an unrestrained way and has been shown to be predictive of personality traits within and outside of selection contexts (e.g., Harrison et al., 2019; Hickman et al., 2021; Hickman et al., 2019; Kern et al., 2019; Liu et al., 2015; Park et al., 2015). For example, research has demonstrated that computational models are able to render fast and relatively accurate predictions of personality from people's social media posts (Kern et al., 2019; Park et al., 2015; Peters & Matz, 2023; Schwartz et al., 2013), predict self- and other-reported personality from applicants' video interviews (Hickman et al., 2022; Hickman et al., 2021; Hickman et al., 2019) or CEOs' personality from transcripts of earning calls (Harrison et al., 2019).

The Present Research

The present research aims to test the performance of ML models in predicting personality traits from text data native to the recruiting process. Focusing on text data generated during the recruiting process makes it possible to leverage the power of predictive models while simultaneously respecting the applicant's privacy and conforming to legal requirements (Goretzko & Israel, 2022; Matz et al., 2020; Matz & Netzer, 2017). We will quantify the predictive performance of our models in two ways. First, the models' construct validity will be measured as the correlation between predicted and self-reported scores. Second, the models' criterion validity will be quantified by correlating the personality predictions with a central external criterion for job matching and preselection (Nye, 2022; Wegmeyer & Speer, 2022): vocational interests (Holland, 1959, 1997). While not as commonly discussed as other constructs (Wegmeyer & Speer, 2022), vocational interests have been shown to be valid predictors of important work-related outcomes such as full-time employment, job stability and income beyond personality and cognitive ability (Stoll et al., 2017). When applied in preselection contexts so that job-interest congruence is achieved, they further have been shown to predict, among others, choice stability (Assouline & Meir, 1987; Hunter & Hunter, 1984), job and job choice satisfaction (Hoff et al., 2020), as well as relevant performance outcomes (Nye et al., 2012, 2017; van Iddekinge et al., 2011), making them a prime construct for preselection and matching.

Building on the existing literature, we pursue the following research questions:

RQ1: How accurately can the Big Five personality traits be predicted from ML-based models using linguistic features extracted from job seekers' CVs and short text responses that were obtained as part of the recruiting process?

RQ2: Can the ML-based personality scores predict vocational interests with a similar accuracy as self-report personality scores?

To address the two research questions, we leverage a combination of open- and closed-vocabulary approaches and train a series of supervised ML models. We compare the performance of models trained on features extracted from (a) applicants' CVs, (b) short free text responses and (c) a combined model with features from both data sources to a baseline model that utilizes information about the applicants' gender. Subsequently, we analyze correlational patterns of self-report and ML-based personality scores with self-report vocational interests.

Methods

Transparency and Openness

We describe our sampling plan, all data exclusions, all manipulations, and all measures in the study. This study was not preregistered. The analysis code is available in the OSF-repository of this project (https://osf.io/gxae9/). Data were analyzed using Python, version 3.7.3 (see respective method section for specific packages). Given that the study relies on secondary data, we did not preregister the design or analyses.

Sample

The data for this study was provided by an intermediary recruiting platform that operates in Australia and aims to place high-performing undergraduate and graduate students in high-quality part-time jobs. To use their services, students are required to set up a general profile that is displayed to possible employers. In the profile set-up process, applicants submit information about themselves including a CV, short textual answers to questions relating to one's planned development trajectory, as well as broadly defined vocational interests. In addition, students are asked to take a short personality test. As part of the sign-up process, students provide consent that their data can be used for research purposes.

The original dataset consisted of 9,280 applicants. For our study, we excluded participants using the following criteria: (i) incomplete personality assessments, (ii) fewer than six words in the short-answer free text responses, and (iii) duplicated applicants. This reduced the original dataset to 8,313 applicants (51.4% male, 47.9% female, 1% other).

Measures and Procedures

CV Data

A total of 7,864 (95%) applicants uploaded documents to the CV section of their profile. Text data was extracted using the *pdfplumber* (version 0.05.28) Python package for all the 7,734 of the pdfs that were machine readable. A manual inspection of the documents revealed that some of the documents contained short test files with only a few words while others included research articles of extreme length. To only capture actual CVs, we excluded all files containing fewer than 50 words or more than 10,000 words. This procedure resulted in a final corpus of 7,691 CV documents with an average number of 561.33 (*SD* = 289.91) words per document.

Free Text Excerpts

As part of the profile creation process, applicants were asked to answer the question "*In a few sentences, please tell us about something you would like to learn more about and why? (150-200 words)"* in written form. On average, applicants in our sample wrote 94.87 words (*SD* = 59.67; see Appendix A for examples).

Big Five Personality Traits

Personality was assessed using the Mini-IPIP (Donnellan et al., 2006), a 20-item version of the 50-item International Personality Item Pool (Goldberg et al., 2006). The Mini-IPIP measures the Big Five with four items per personality trait. Participants were asked to rate the items on a 5-point Likert scale (0 = "Strongly disagree", 4 = "Strongly agree"). Table 1 displays means, standard deviations, and internal consistencies.

Table 1.
Cronbach's Alpha, Average Scores and Standard Deviation of the Big Five Personality Traits

Personality Trait	Cronbach's Alpha	M	SD
Openness	.68 (.65)	3.12	0.57
Conscientiousness	.62 (.69)	3.01	0.59
Extraversion	.78 (.77)	2.57	0.74
Agreeableness	.67 (.70)	3.30	0.51
Neuroticism	.57 (.69)	1.25	0.61

Note. N = 8,313. Values in parentheses display values from the original validation study (Donnellan et al., 2006). *M* = sample's averaged scores per personality dimension. *SD* = samples standard deviation per personality dimension.

Vocational Interests

Vocational interest data was available for a subset of n = 3,469 of applicants. Applicants were asked to select their interests from a set of eight pre-defined categories using a "tick-all-that-apply" format: Design, marketing, programming, finance, analytics, operations, accounting, and HR. On average, applicants selected 3.46 categories (SD = 1.65; see Appendix Table B1 for choice frequencies across all categories).

Prediction Models

The development and validation of our prediction models was based on two major steps: (1) Preprocessing and feature extraction and (2) training and validation. Figure 1 illustrates the process visually.

Step 1: Preprocessing of Text Data and Feature Extraction

Our text-based models extended the feature extraction process by Park et al. (2014) by combining closed- with open-vocabulary approaches. Specifically, we extracted four types of language features: (1) words and phrases, (2) topics, (3) LIWC-features and (4) word embeddings. In

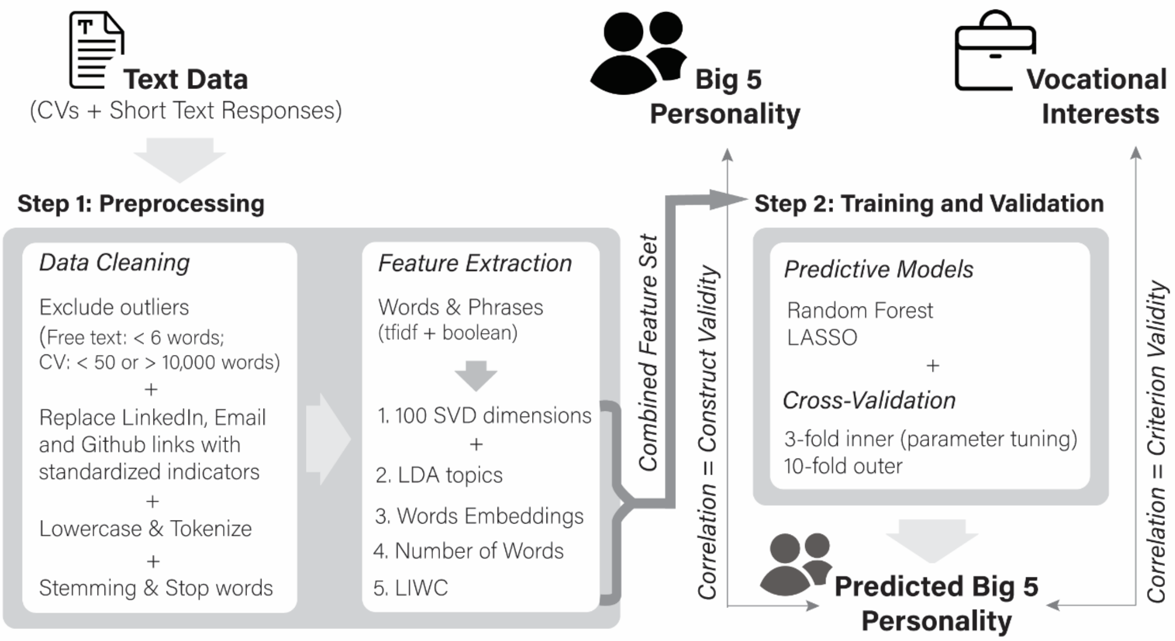

Figure 1. Process flow diagram illustrating the modeling approach used to build the ML-based models predicting personality.

addition, we also included the length of the answers as a feature. The final feature set included 694 predictors.

For the purpose of feature extraction, we transformed all text input to lowercase and split the text into tokens (i.e., single words). Afterwards, we removed stop words (i.e., very frequent words like *a*, *the,* or *is*) and stemmed the words using the Porter stemmer integrated in the Python package nltk (version 3.4.5; Bird et al., 2009). For the CVs, we replaced references to LinkedIn profiles, email addresses and GitHub repositories with the tokens “LINKEDIN”, “EMAILADDRESS” and “GITHUB” to allow them to be included as features in the frequency-based feature extraction methods. For the feature extraction using LIWC, the unprocessed answers were directly fed into the LIWC software, which also examines linguistic styles that are extracted based on stopwords (Pennebaker et al., 2015).

Word Frequencies (Count Vectorizer, Tf-idf). After preprocessing, we first extracted word frequencies and short phrases using scikit-learn (version 0.21.2; Pedregosa et al., 2011). Phrases refer to combinations of words that occur next to each other (n-grams). Parallel to Park et al. (2015) we included phrases up to three words (e.g. “My name is”). These frequencies were then weighted based on their term frequency-inverse document frequency (tf-idf). Tf-idf weighting can be used to determine how important a word is in a document that is part of a larger corpus. Tf-idf values are high for words with a high term frequency (i.e., occur often in a specific document) and a low document frequency in the whole corpus (i.e., does not occur often in general). In addition to these tf-idf weighted frequencies, we added a binary representation of each of the extracted frequency features indicating if an applicant used a word or a phrase at least once in his answer. Since the number of features generated by considering words and word combinations reached into the hundred thousands, we decided to limit the number of extracted features to 10,000 (5,000 tf-idf weighted features and 5,000 binary features). These two sets of features were extracted by each considering the top 5,000 features ordered by term frequency. Given the limited sample size of our data, we subsequently applied Singular Value Decomposition, a method for dimensionality reduction

(Golub & Reinsch, 1970) to each of the feature sets of 5,000 features reducing each of them to 100 features.

Topics. The second feature set we extracted is based on topics. Topics refer to clusters of semantically related words that are generated through Latent Dirichlet Allocation (LDA; Blei et al., 2003). We fitted the LDA model using the LatentDirichletAllocation function implemented via scikit-learn (version 0.21.2; Pedregosa et al., 2011), setting the number of topics to 100. This created a document-topic matrix with documents as rows and the usage of each topic per document as columns, resulting in a feature set of 100 topic features.

LIWC. LIWC analyzes text data by first comparing the words of a text against predefined categories (e.g., the word "hate" is associated with the categories Affective Processes, Negative Emotion and Anger) and then creating category-based frequency scores. Based on this procedure, LIWC creates a matrix of 93 output variables, including single linguistic styles and broader summary variables, which we all included in our feature set. For a complete list of the LIWC2015 scales see the official LIWC2015 language manual (Pennebaker et al., 2015).

Word Embeddings. Embedding methods convert symbolic, abstract representations of meaning such as words and images into numeric vector representations. These vectors capture the underlying semantic relations between the symbolic representations. The intuition behind word embeddings is reflected by the idea that the meaning of a word can be deducted by analyzing the sets of words that commonly surround it across many different contexts. The result is an n-dimensional feature space representing the meaning of word in relation to other words (Lawson et al., 2022).

We used the pre-trained 300-dimensional vector package "en_core_web_md" from Stanford's GloVe Project (Pennington et al., 2014) which is implemented in the Python package spaCy (Version 2.0.12; Montani et al., 2022) and includes vector representations for 20,000 different words and accounts for conjunctions of words by mapping its vectors against 650,000 keys. We implemented these pre-trained embeddings by making use of spaCy's function to create document vectors. Document vectors are created by first extracting vector representations for the single words in a document and then calculating the mean vector for each of the 300 dimensions over all words in a document. The resulting feature set is a 300-dimensional set of vectors with one 300-dimensional vector for each document.

Step 2: Model Training and Validation

The combined feature set of 694 features was used to train a Random Forest regression model predicting personality scores. The Random Forest algorithm is an ensemble machine learning technique that models non-linear relationships (Breiman, 2001). The algorithm simultaneously constructs multiple decision trees by randomly drawing subsets of input data (features) and then aggregating these trees to improve predictive accuracy. We trained a total of three models: First, the free text-model, trained on the 694 features extracted from the free texts. Second, the CV-model, trained on the 694 features from the CVs. And third, a combined model, trained on a stacked feature set including both the features extracted from the free texts and the CVs with 1,388 features. To provide an additional reference point against which our language-based personality predictions could be evaluated, we also trained a linear regression model using gender as the sole predictor (baseline model).

To analyze the robustness of our findings we additionally performed the same analyses with linear Lasso models (Tibshirani, 1996, see Appendix C for more details). To keep the discussion of our findings concise, we only report the findings for the Random Forest models in the main manuscript and report the findings for the Lasso models in the Supplementary Information. As expected, the more complex non-linear Random Forest models outperformed the Lasso models, but notably the difference in accuracy was found to be relatively small suggesting that even simple linear models capture a considerable amount of variance in personality traits. All algorithms were implemented using the scikit-learn package in Python (version 0.21.2; Pedregosa et al., 2011).

To train and evaluate our models in an iterative process, we applied a nested resampling strategy (Stachl et al., 2020). This strategy nests multiple repeating resampling loops in each other, separating the fitting process and the hyperparameter tuning process. We used two nested loops: An inner loop for hyperparameter tuning in which we applied 3-fold cross validation and an outer loop for the fitting of the models using the selected hyperparameters in which we applied 10-fold cross-validation. For the hyperparameter tuning of the Random Forest models we implemented a random search with the default configuration of 100 iterations. For the Lasso models we used a grid search to find the optimal value for the penalization parameter λ (see Appendix Table C1 for hyperparameter ranges). Performance was evaluated by correlating the self-report personality scores with the predicted personality scores from the outer loop.

Results

Construct Validity: Correlations Between Predicted and Self-Reported Personality Traits

Figure 2 displays the Pearson correlations between the predicted and self-report personality scores for the Random Forest models (see Appendix Table D1 for a table of Random Forests results and Appendix Table D2 for the results from the Lasso models). All three text-based Random Forest models showed accuracies that were significantly different from zero and outperformed the baseline model (average correlation of baseline *r* = .06). Averaged across the five traits, personality could be

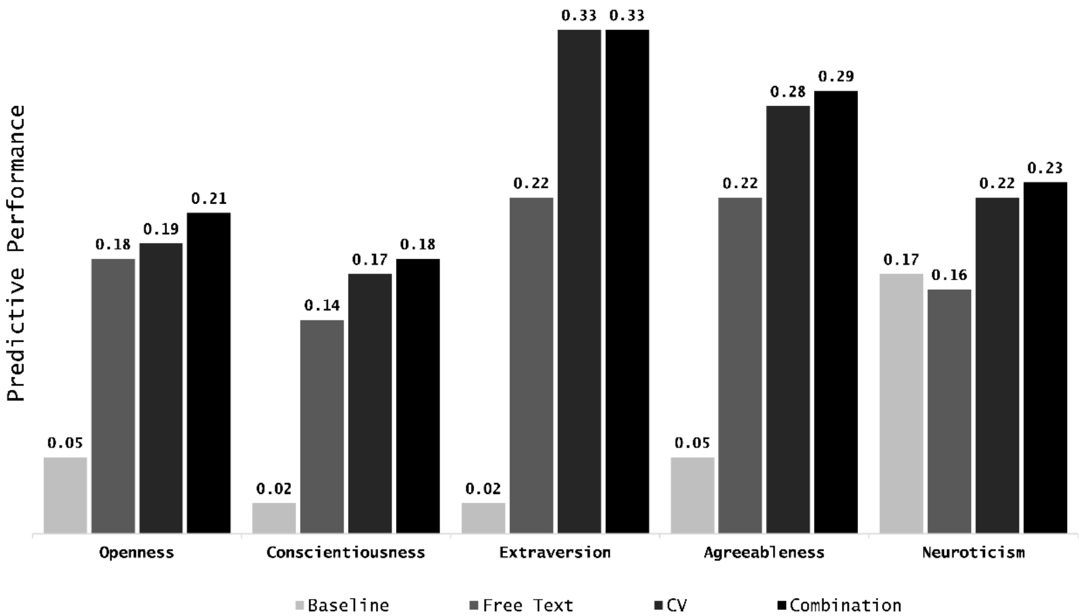

Figure 2. Bar charts comparing the predictive performance of the different Random Forest models per Big Five personality trait. Predictive performance is defined as correlation between self-report and predicted personality. From left to right within one trait: (1) Baseline model, (2) free text model, (3) CV model, (4) combined model.

predicted with an accuracy of *r* = .18 from the free text data, *r* = .24 from the CV data and *r* = .25 from the combination of both data sources. For all three text-based models, the highest performance was achieved for the dimension of Extraversion (free text: *r* = .22, CV: *r* = .33, combined: *r* = .33).

To benchmark the predictive accuracies of our models to current best practices, we compared them to the accuracy of human judges (see Figure 3). Prior research suggests that recruiters regularly infer personality from CVs, but do so inaccurately (Apers & Derous, 2017; Burns et al., 2014; Cole et al., 2009). Cole et al. (2009) found that when presented with written CVs, recruiters were only able to judge Extraversion better than chance (*r* = .15). A similar pattern was replicated in research on personality judgments based on LinkedIn profiles that include both free text sections and CVs. Both, Roulin & Levashina (2019) and Van de Ven et al. (2017) found that judges were only able to judge Extraversion better than chance (Roulin & Levashina, 2019: *r* = .20; van de

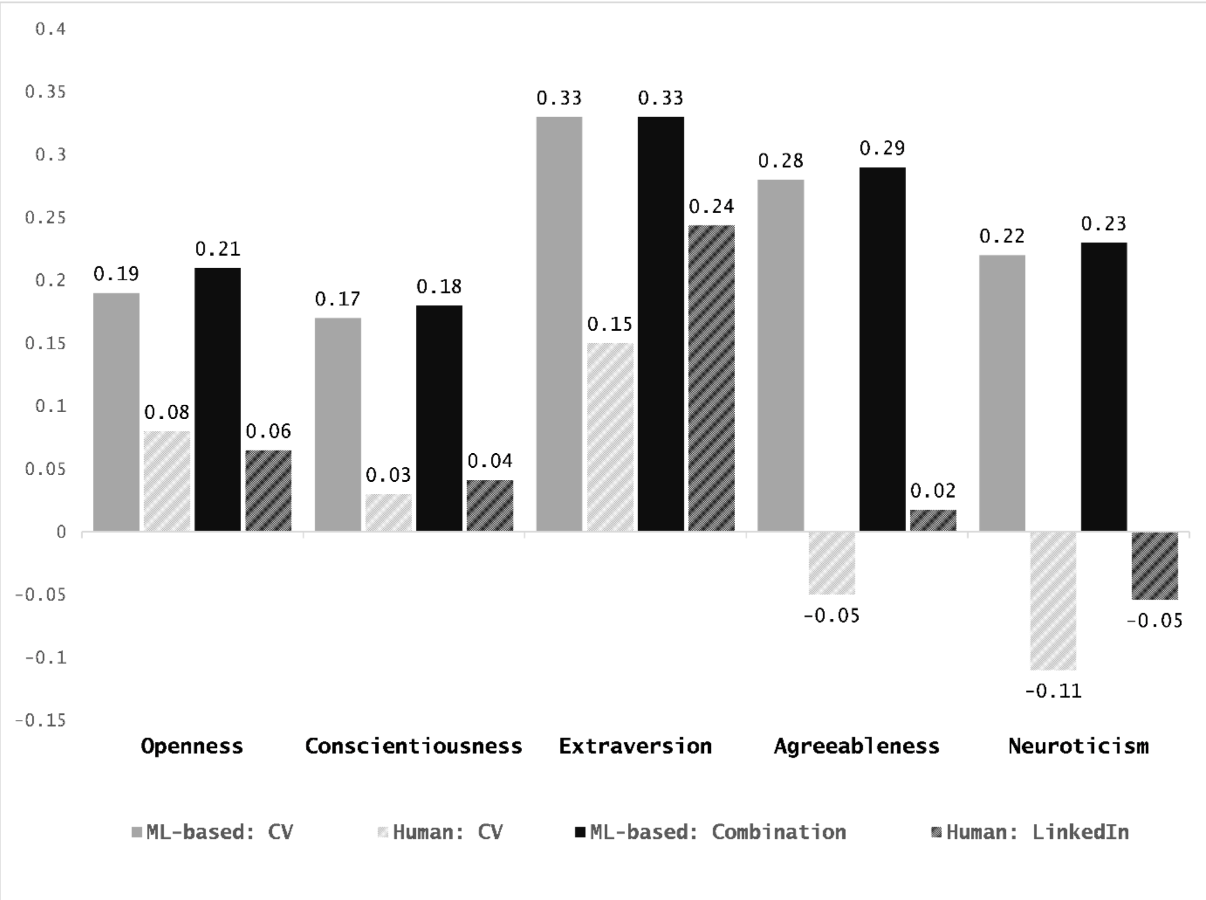

Figure 3. Bar charts comparing the predictive performance of the Random Forest models against the performance of human raters on similar material reported in prior research. Predictive performance is defined as correlation between self-report and predicted personality. From left to right within one personality trait: (1) ML-based CV model, (2) human performance rating personality from CVs (Cole et al., 2009), (3) combined ML-based model, (4) human performance rating personality from LinkedIn profiles (calculated as weighted mean from results from Roulin & Levashina, 2019 and van de Ven et al., 2017).

Ven et al., 2017: $r = .29$). Our models outperform these accuracies of human judges by (a) making better than chance predictions for all five personality dimensions, and surpassing the highest reported average human accuracy both (b) across traits ($r = .08$; Roulin & Levashina, 2019) and (c) for a single trait dimension (Extraversion $r = .29$; van de Ven et al., 2017).

Criterion Validity: Correlations Between Predicted Personality Scores and Vocational Interests

Table 2 displays correlations between vocational interests and the four assessment approaches to personality: Self-report personality, the free text Random Forest model, the CV Random Forest model, and the combined Random Forest model. Supporting the validity of our predictive models, all text-based models showed similar correlational patterns with the vocational interests as the self-reported personality traits, with no difference in direction of any of the significant correlations. In fact, the predicted scores of all computational models showed higher average correlations with vocational interests (free text $r = .09$; CV and combined model $r = .14$) than the self-report personality scores ($r = .05$). See Appendix Table D3 for the respective Lasso results.

Figure 4 shows that the vast majority of correlations between the five personality traits and eight vocational interests (90%) are stronger for ML-based than self-reported personality traits (i.e., observations located in the upper diagonal).

.

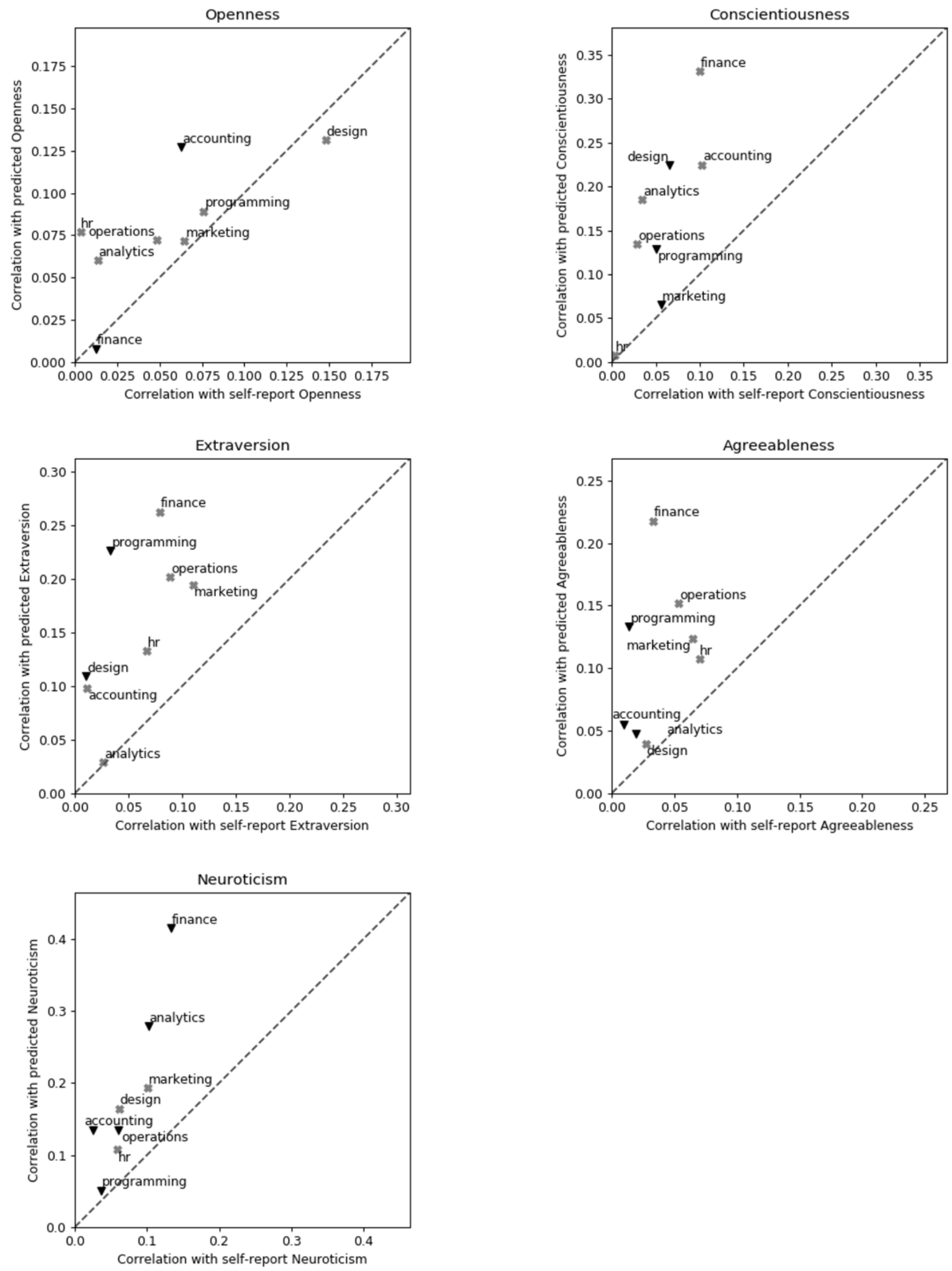

Figure 4. Scatterplots displaying absolute correlation between predicted and self-report personality scores and self-report vocational interest categories for the combined Random Forest model. Dots positioned above the diagonal line display cases in which the ML-based personality traits correlated stronger with vocational interests than the self-report models. Positive correlations are displayed as grey “X”s and negative correlations are displayed as black triangles. Differences in direction of the correlations between predicted and self-report personality scores only occurred with non-significant correlations. In these cases, the displayed mark represents the direction of the correlation of the predicted personality scores with the self-report vocational interest categories.

Table 2

Correlation of the Big Five Self-Report Personality Traits and Random Forest ML-Based Models with Indications of Vocational Interests

	Self-report					Free text					CVs					Combined				
	O	C	E	A	N	O	C	E	A	N	O	C	E	A	N	O	C	E	A	N
Des	**.15**	**-.07**	-.01	.03	**.06**	**.11**	**-.16**	**-.08**	-.01	**.14**	**.14**	**-.20**	**-.11**	**-.04**	**.16**	**.13**	**-.22**	**-.11**	**-.04**	**.16**
Mar	**.06**	**-.06**	**.11**	**.06**	**.10**	.01	**-.05**	**.09**	**.05**	**.15**	**.08**	**-.08**	**.19**	**.14**	**.19**	**.07**	**-.07**	**.19**	**.12**	**.19**
Pro	**.08**	**-.05**	-.03	-.01	**-.04**	**.09**	**-.08**	**-.11**	**-.04**	**-.05**	**.08**	**-.12**	**-.21**	**-.14**	**-.04**	**.09**	**-.13**	**-.23**	**-.13**	**-.05**
Fin	-.01	**.10**	**.08**	.03	**-.13**	.01	**.21**	**.16**	**.11**	**-.24**	.02	**.34**	**.26**	**.21**	**-.41**	.01	**.33**	**.26**	**.22**	**-.42**
Ana	.01	**.03**	.03	-.02	**-.10**	**.04**	**.13**	.03	**.04**	**-.20**	**.05**	**.19**	.03	.03	**-.26**	**.06**	**.18**	.03	**.05**	**-.28**
Ope	**.05**	.03	**.09**	**.05**	**-.06**	**.04**	**.08**	**.11**	**.10**	**-.08**	**.08**	**.13**	**.19**	**.14**	**-.13**	**.07**	**.13**	**.20**	**.15**	**-.13**
Acc	**-.06**	**.10**	.01	-.01	-.02	**-.09**	**.14**	**.06**	.01	**-.08**	**-.11**	**.23**	**.09**	**.05**	**-.13**	**-.13**	**.22**	**.10**	**.06**	**-.13**
HR	.00	.00	**.07**	**.07**	**.06**	-.03	-.02	**.10**	**.07**	**.08**	**-.09**	.01	**.13**	**.12**	**.10**	**-.08**	-.01	**.13**	**.11**	**.11**
M	**.05**	**.06**	**.05**	**.04**	**.07**	**.05**	**.11**	**.09**	**.05**	**.13**	**.08**	**.16**	**.15**	**.11**	**.18**	**.08**	**.16**	**.16**	**.11**	**.19**

Note. N = 3,469. M = Average correlation within each column which was calculated by first extracting absolute values, applying Fisher's r-to-z transformation to each correlation, averaging, and then transforming back to r. *O* = Openness, *C* = Conscientiousness, *A* = Agreeableness, *N* = Neuroticism. *Des* = Design, *Mar* = Marketing, *Pro* = Programming, *Fin* = Finance, *Ana* = Analytics, *Ope* = Operations, *Acc* = Accounting.

Bold = p < .05.

Discussion

This paper set out to address two critical questions related to the feasibility and value of using ML-based predictions of personality in preselection and job matching contexts: How accurately can personality be predicted from textual information obtained early and solely in the recruiting process (construct validity)? And how do the predicted personality scores relate to relevant external criteria, such as applicants' vocational interests (criterion validity)?

Providing support for the construct validity of the suggested approach (RQ1), our findings suggest that applicants' personality traits can be predicted from text data obtained early in the recruiting process. All three text-based models – free text response, CV and combined – showed significant correlations and outperformed the demographic baseline model. The highest average performance was obtained by the combined model (r = .25), followed by the CV model that is just marginally less performant (r = .24) but stronger than the free text model (r = .18).

While the accuracies obtained by our text-based models are on par with or superior to those observed for other digital footprints, including music preferences (average r = .17; Nave et al., 2018), Flickr Pictures (average r = .18; Segalin et al., 2017) and spending records (average r = .21; Gladstone et al., 2019), as well as specific language cues from hiring interviews (average r = .19; Hickman et al., 2021), they are lower than those observed for social media data, including both Facebook Likes (r = .45; Youyou et al., 2015) and Facebook status updates (r = .41; Park et al., 2015). There are multiple explanations for this finding. First, the accuracy of computational models depends on the amount and richness of the data. While our sample of 8,313 applicants and around 660 words per applicant is unique, it is still considerably smaller than the Facebook samples which included data from more than 71,000 users with an average of about 4,100 words per user (e.g., Park et al., 2015). Second, there are important differences in the type of information used. Facebook statuses consist of information people intentionally share with other people and therefore reflect content such as opinions and emotions. In contrast, in a recruiting setting, applicants (a) are prompted to respond to a specific question and submit specific material (e.g., CVs) and (b) know that their information will be reviewed with regard to a hiring decision. Thus, their answers are more restricted and do not allow for a free expression of their personality.

Notably, our models still substantially outperformed accuracies that have been reported for human judges (Apers & Derous, 2017; Cole et al., 2009; Roulin & Levashina, 2019; van de Ven et al., 2017). This comparison not only has practical implications, but also shows that relevant cues for personality judgments are available in application materials. Whereas human judges are known to be able to judge the Big Five personality traits of strangers from cues found in physical and online spaces (e.g., bedrooms or online social media profiles; Back et al., 2010; Gosling et al., 2002; Küfner et al., 2010; Naumann et al., 2009), prior work has suggested that recruiters struggle to make valid personality judgments based on application material for all Big 5 traits but for Extraversion (Apers & Derous, 2017; Cole et al., 2009; Roulin & Levashina, 2019; van de Ven et al., 2017).

One possible explanation for this finding is that the application context is characterized by relatively strict and widely understood norms about the structure and content of its materials (Apers & Derous, 2017). These norms and expectations restrict the richness of expression in the application documents and reduce the availability of relevant cues for personality judgments. For instance, a possible explanation for the challenge of human judges assessing Conscientiousness in application materials (Apers & Derous, 2017; Cole et al., 2009; Roulin & Levashina, 2019; van de Ven et al., 2017) compared to social media profiles (Back et al., 2010; Hall et al., 2014; Marcus et al., 2006) is that, unlike social media's more socially focused self-presentation contexts, the expectation of neatness and accuracy in application materials may prompt candidates to engage in more extensive error-checking behavior. This behavior then minimizes the variance in otherwise overtly available cues for Conscientiousness such as the existence of spelling or grammatical errors. However, the fact that our models were able to predict personality with some degree of accuracy suggests that application materials indeed contain personality cues that might be overlooked by human judges. This finding underscores the potential of ML methods to discover patterns in complex data sources and to complement human workers in their day-to-day tasks.

Providing support for the criterion validity of our text-based models (RQ2), the predicted personality scores showed consistent correlational patterns with vocational interests that were aligned with the correlations observed for the self-reported personality scores (see Table 2). What is more, the observed correlations of the predicted personality scores with the vocational interests were found to be higher (average r = .15) than the self-report personality scores (average r = .05). These findings are aligned with the results of prior research finding that predicted personality scores correlate more strongly with certain external variables than self-report personality (Park et al., 2015; Youyou et al., 2015). A potential explanation for this phenomenon is that the predictive models might only capture a specific part of the variance in the outcome (i.e., self-report personality scores) that is pertinent to the training data. First, the applied ML models inherently aim to capture stable patterns in the features that are predictive of the outcome. As noise in both, the features and the outcome, should not be stable but random, the resulting predictions should not include some of the noise inherent to the variance of the self-report personality scores. Second, since the models were trained on features extracted from application materials, the predicted personality score might capture individual differences that are more directly work-related and therefore show stronger associations with vocational interests than self-report personality scores. If this trend can be replicated using different types of training data, the use of predicted personality scores may not only add value by being more efficient and scalable, but also by representing less noisy and more domain-relevant constructs than broad self-report personality scores.

Practical Implications

Our findings showcase that psychological characteristics and preferences of job candidates can be predicted using readily available data native to the recruiting process early on without the need to access external third-party information. Insights into these characteristics are critical when it comes to matching the right candidate to the right job, thereby reducing the risk of quick turnover, and lowering overall hiring costs. In contrast to previous work that relies on social media data to predict psychological characteristics (e.g., Park et al., 2015; Youyou et al., 2015), our approach offers several advantages with regard to privacy and legal concerns. Because our models exclusively rely on information that was voluntarily shared by participants in the context of the recruiting process for the purpose of evaluating their fit for a given job, they are aligned with recent conceptualizations of privacy as contextual integrity (Nissenbaum, 2004): A person's privacy is upheld if the flow and use of data is aligned with the expectations that the individual has about how their data is being used.

Importantly, we do not mean to overstate the accuracies of our predictive models. While our findings suggest that ML-based predictions are likely to outperform the judgements made by human managers and are superior to self-reports in predicting vocational interests, the absolute accuracies are still moderate at best. Given the importance of hiring decisions for both companies and applicants, we therefore caution against the use of such predictive models as final decision-making tools. Instead, they could be used to assists managers in asking the right questions during the interview stage, or to proactively recommend and preselect relevant job openings to candidates. This approach would be aligned with recent updates to privacy legislation such as the General Data Protection Regulation (GDPR), which restrict the use of profiling when it comes to making automated decisions on the bases of such profiles (European Parliament & European Council, 2016).

Limitations, Constraints on Generality & Future Research

The present study comes with a number of limitations that should be addressed by future research. First, the data for this study stems from a highly selective sample. It was collected as part of a recruiting process aimed at placing high-performing students into high-quality student jobs (e.g., student jobs at consulting firms, IT firms, banks). As a case in point, this restriction resulted in smaller standard deviations in personality scores compared to the original validation study of the Mini-IPIP (Donnellan et al., 2006). While this restriction suggests that the predictive performance of ML-based models was conservatively estimated in the current study and might be higher in more heterogeneous samples, future research should address this question on the generality of our results empirically. What is more, future research should explore potential moderators of the models' predictive performance. For example, future work could investigate the effect of sample size and amount of information per applicant on predictive performance.

Second, to ensure the fairness of predictions of our models, future work should investigate whether factors inherent to the person might influence predictive accuracy and potentially lead to algorithmic bias (Cowgill & Tucker, 2020). Empirically this bias could manifest in different ways. First, the predictive models might predict different score levels of personality traits for different subgroups (e.g., native vs. non-native speakers, race, gender) despite them sharing the same ground-truth levels for the respective personality construct. Second, the models might produce different levels of predictive accuracy across subgroups (Tay et al., 2022). For example, it is possible that the accuracy of our text-based models is lower for non-native speakers who do not have access to the full vocabulary of a native speaker and might therefore find it harder to express their personalities. This can be true even when both groups share the same levels for the latent personality constructs. To uncover such biases, future research should first identify relevant subgroups and then perform subgroup analyses investigating whether differences exist in the distribution of the predicted personality scores or the models’ accuracy across these subgroups.

Moreover, future research should implement interpretable ML approaches to identify potential biases and develop a better understanding of which aspects of application materials are predictive of certain personality judgements (Molnar, 2022). In our case, the relatively small dataset made it necessary to reduce the complexity of the text data by applying a dimensionality reduction step via SVD. However, larger datasets would offer the option to process the text data at the level of n-grams, therefore retaining the ability to meaningfully interpret model outputs. One option to integrate an interpretable ML approach into the modeling pipeline would be to include permutation-based feature importance measures (Breiman, 2001) into the outer-loop of the cross-validation approach. These feature importance metrics are calculated by evaluating the change in predictive performance subsequent to random permutation (i.e., shuffling) of the values of a feature. This procedure makes it possible to determine how important each feature is for the predictive performance of the model. In a first step, the extracted feature importance metrics could then be used to understand how the models integrate the extracted language feature into a prediction. In a second step, they could also be used to analyze whether the information the model relies on when making its predictions varies across different subgroups (e.g., gender, native speaker).

An alternative approach could be the extraction of SHAP values (Lundberg & Lee, 2017). SHAP values are rooted in cooperative game theory and, while generally being computationally more expensive, come with the advantage that they allow for global and local explanation. That is, whereas permutation feature importance measures are usually used to explain the decision rules of a model across all of its predictions (global explanation), SHAP values are designed to also provide explanations for every individual prediction (i.e., every applicant in our dataset; local explanation). Having access to local explanations might bring additional benefits in the context of recruiting as recruiters could see how the information from the application material was used in the personality prediction for specific applicants. Having access to local explanations could help to identify potential biases in the models at a more fine-grained level. Equally, local explanations could also be used to transparently demonstrate that applicants have been treated fairly and that the predictions were not driven by biases if for example mandated in a lawsuit. As of today, these bias analyses and more would potentially also be required to adhere to the proposed EU AI Act (AI Act, 2021). Exact assessments are not possible as the AI Act to this date is still under discussion and has not yet been passed into legislature.

Third, this study only considered two types of text data: free text responses and CVs. Future research should investigate how different types of text data from the recruiting process such as cover letters or letters of reference can be leveraged in prediction models. An intriguing question is how recruiting processes could be designed to produce even richer, more predictive data sources, for example by comparing different questions, asking for more free text answers or requiring applicants to produce more text in conversations with chat bots.

Fourth, in line with prior research we found that the predicted personality values correlate higher with certain external variables than self-report personality (Park et al., 2015; Youyou et al., 2015). An investigation of the unique contributions of ML-based personality models over self-report assessment, especially as a function of their training data, could (a) help map the advantages of the predicted constructs over self-report-assessed constructs and (b) uncover possible blind-spots of self-reports. To further investigate this phenomenon, research could map predicted personality scores from models trained on different types of training data and their self-report counterparts to different external criteria. For example, for the case of predicted personality scores in recruiting settings, predicted personality scores based on work-related and work-unrelated data sources should be mapped to a variety of different external criteria such as cognitive ability, job satisfaction or job performance.

Fifth, we applied a combined approach of open- and closed-vocabulary extraction approaches. Since the field of natural language processing is constantly evolving, new methods such as contextualized word embeddings like BERT (Devlin et al., 2018) or more recently large language models (Touvron et al., 2023) become available to researchers at increasingly narrow intervals. Future research should investigate how models can be improved by including these new methods.

Sixth, to better understand the practical implications of our results, future research should examine the downstream value of implementing ML-based personality prediction models in real-world recruiting contexts, for example, in a recommendation system that proactively matches candidates to jobs. Are applicants matched using ML-based predictions of personality more likely to receive an initial response from the employer? Are they more likely to be hired for the job after going through the full interview process? Are they happier and more productive in their job? And are they less likely to leave? Providing answers to these questions will be critical for showing the real value of our predictive models.

Conclusion

In this paper, we demonstrated the feasibility of using automated predictions of personality in applied preselection and recruiting settings. We extended existing research by showing that (i) personality traits can be predicted from text data generated early in the recruiting process (rather than later or from external third-party data, such as social media profiles) and that (ii) such automated personality assessments are able to predict important external criteria such as vocational interests. Our findings highlight an opportunity for organizations to utilize automated assessments of personality in their recruitment processes without running the risk of violating participants' privacy and encountering legal challenges. Based on the observed accuracies of our models, we caution against the use of such predictive models to make final hiring decisions, but instead encourage their application as managerial decision-making aids and input into proactive job recommendations.

References

Acemoglu, D., & Restrepo, P. (2017). Robots and jobs: Evidence from US labor markets: NBER Working Paper No. 23285. Advance online publication. https://doi.org/10.3386/w23285

Acemoglu, D., & Restrepo, P. (2019). Artificial intelligence, automation, and work. In A. Agrawal, A. Goldfarb, & J. Gans (Eds.), *National Bureau of Economic Research conference report. The economics of artificial intelligence: An agenda* (pp. 197–236). The University of Chicago Press. https://doi.org/10.7208/chicago/9780226613475.001.0001

Allen, D. G., Mahto, R. V., & Otondo, R. F. (2007). Web-based recruitment: Effects of information, organizational brand, and attitudes toward a web site on applicant attraction. *Journal of Applied Psychology*, *92*(6), 1696–1708. https://doi.org/10.1037/0021-9010.92.6.1696

Anderson, H. J., Baur, J. E., Griffith, J. A., & Buckley, M. R. (2017). What works for you may not work for (Gen)Me: Limitations of present leadership theories for the new generation. *The Leadership Quarterly*, *28*(1), 245–260. https://doi.org/10.1016/j.leaqua.2016.08.001

Apers, C., & Derous, E. (2017). Are they accurate? Recruiters' personality judgments in paper versus video resumes. *Computers in Human Behavior, 73*, 9–19. https://doi.org/10.1016/j.chb.2017.02.063

Assouline, M., & Meir, E. I. (1987). Meta-analysis of the relationship between congruence and well-being measures. *Journal of Vocational Behavior*, *31*(3), 319–332. https://doi.org/10.1016/0001-8791(87)90046-7

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

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*(3), 372–374. https://doi.org/10.1177/0956797609360756

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., & Mount, M. K. (2012). Nature and use of personality in selection. In N. Schmitt (Ed.), *Oxford Handbook of Personnel Assessment and Selection* (pp. 225–251). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199732579.013.0011

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*(1&2), 9–30. https://doi.org/10.1111/1468-2389.00160

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, *92*(3), 595–615. https://doi.org/10.1037/0021-9010.92.3.595

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*(2), 410–424. https://doi.org/10.1037/0021-9010.92.2.410

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python* (1st). O'Reilly Media. https://doi.org/10.5555/944919.944937

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022. https://doi.org/10.5555/944919.944937

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, *89*(5), 901–910. https://doi.org/10.1037/0021-9010.89.5.901

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burns, G. N., Christiansen, N. D., Morris, M. B., Periard, D. A., & Coaster, J. A. (2014). Effects of applicant personality on resume evaluations. *Journal of Business and Psychology*, *29*(4), 573–591. https://doi.org/10.1007/s10869-014-9349-6

Cardoso, A., Mourão, F., & Rocha, L. (2021). The matching scarcity problem: When recommenders do not connect the edges in recruitment services. *Expert Systems with Applications*, *175*, 114764. https://doi.org/10.1016/j.eswa.2021.114764

Cole, M. S., Feild, H. S., Giles, W. F., & Harris, S. G. (2009). Recruiters' inferences of applicant personality based on resume screening: Do paper people have a personality? *Journal of Business and Psychology*, *24*(1), 5–18. https://doi.org/10.1007/s10869-008-9086-9

Cowgill, B., & Tucker, C. E. (2020). Algorithmic fairness and economics. *Columbia Business School Research Paper.* Advance online publication. https://doi.org/10.2139/ssrn.3361280

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018, October 11). *BERT: Pre-training of deep bidirectional transformers for language understanding.* https://doi.org/10.18653/v1/N19-1423

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and ammending certain union legeslative acts, April 21, 2021.

European Parliament, & European Council (2016). EU General Data Protection Regulation (EU-GDPR). *Official Journal of the European Union*, *119*(1).

Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality*, *43*(2), 120–126. https://doi.org/10.1016/j.jrp.2008.12.041

Gladstone, J. J., Matz, S. C., & Lemaire, A. (2019). Can psychological traits be inferred from spending? Evidence from transaction data. *Psychological Science*, *30*(7), 1087–1096. https://doi.org/10.1177/0956797619849435

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037/0003-066x.48.1.26

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numerische Mathematik*, *14*(5), 403–420. https://doi.org/10.1007/BF02163027

Goretzko, D., & Israel, L. S. F. (2022). Pitfalls of machine learning-based personnel selection: Fairness, transparency, and data quality. *Journal of Personnel Psychology*, *21*(1), 37–47. https://doi.org/10.1027/1866-5888/a000287

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*(3), 379–398. https://doi.org/10.1037//0022-3514.82.3.379

Hall, J. A., Pennington, N., & Lueders, A. (2014). Impression management and formation on Facebook: A lens model approach. *New Media & Society*, *16*(6), 958–982. https://doi.org/10.1177/1461444813495166

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal.* Advance online publication. https://doi.org/10.1002/smj.3023

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*(3), 639–683. https://doi.org/10.1111/j.1744-6570.2004.00003.x

Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2022). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology*, *107*(8), 1323–1351. https://doi.org/10.1037/apl0000695

Hickman, L., Saef, R., Ng, V., Woo, S. E., Tay, L., & Bosch, N. (2021). Developing and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews. *Human Resource Management Journal.* Advance online publication. https://doi.org/10.1111/1748-8583.12356

Hickman, L., Tay, L., & Woo, S. E. (2019). Validity evidence for off-the-shelf language-based personality assessment using video interviews: Convergent and discriminant relationships with self and observer ratings. *Personnel Assessment and Decisions*, *5*(3). https://doi.org/10.25035/pad.2019.03.003

Hoff, K. A., Song, Q. C., Wee, C. J., Phan, W. M. J., & Rounds, J. (2020). Interest fit and job satisfaction: A systematic review and meta-analysis. *Journal of Vocational Behavior*, *123*, 103503. https://doi.org/10.1016/j.jvb.2020.103503

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, *6*(1), 35–45. https://doi.org/10.1037/h0040767

Holland, J. L. (1997). *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments* (3rd ed.). Psychological Assessment Resources.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. https://doi.org/10.1037/0033-2909.96.1.72

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765–780. https://doi.org/10.1037//0021-9010.87.4.765

Kern, M. L., McCarthy, P. X., Chakrabarty, D., & Rizoiu, M.-A. (2019). Social media-predicted personality traits and values can help match people to their ideal jobs. *Proceedings of the National Academy of Sciences*, *116*(52), 26459–26464. https://doi.org/10.1073/pnas.1917942116

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. J. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *The American Psychologist*, *70*(6), 543–556. https://doi.org/10.1037/a0039210

Kosinski, M., Stillwell, D. J., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*(1), 1–49. https://doi.org/10.1111/j.1744-6570.1996.tb01790.x

Kristof-Brown, A., & Guay, R. P. (2011). Person–environment fit. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Vol. 3: Maintaining, expanding, and contracting the organization* (pp. 3–50). American Psychological Association. https://doi.org/10.1037/12171-001

Küfner, A. C., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality*, *44*(4), 427–435. https://doi.org/10.1016/j.jrp.2010.05.003

Lawson, M. A., Martin, A. E., Huda, I., & Matz, S. C. (2022). Hiring women into senior leadership positions is associated with a reduction in gender stereotypes in organizational language. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(9), 1–11. https://doi.org/10.1073/pnas.2026443119

Liu, P., Tov, W., Kosinski, M., Stillwell, D. J., & Qiu, L. (2015). Do Facebook status updates reflect subjective well-being? *Cyberpsychology, Behavior and Social Networking*, *18*(7), 373–379. https://doi.org/10.1089/cyber.2015.0022

Lub, X. D., Bal, P. M., Blomme, R. J., & Schalk, R. (2016). One job, one deal…or not: Do generations respond differently to psychological contract fulfillment? *The International Journal of Human Resource Management*, *27*(6), 653–680. https://doi.org/10.1080/09585192.2015.1035304

Lund, S., Manyka James, & Robinson Kelsey. (2016, March 1). *Managing Talent in a Digital Age* (McKinseyQuarterly). McKinsey Global Institute.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems.* Curran Associates, Inc. http://arxiv.org/pdf/1705.07874v2

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, *90*, 46–60. https://doi.org/10.1016/j.futures.2017.03.006

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., & Sanghvi, S. (2017). *Jobs lost, jobs gained: Workforce transitions in a time of automation.* McKinsey Global Institute.

Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology*, *90*(6), 1014–1031. https://doi.org/10.1037/0022-3514.90.6.1014

Matz, S. C., Appel, R. E., & Kosinski, M. (2020). Privacy in the age of psychological targeting. *Current Opinion in Psychology*, *31*, 116–121. https://doi.org/10.1016/j.copsyc.2019.08.010

Matz, S. C., & Netzer, O. (2017). Using Big Data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences*, *18*, 7–12. https://doi.org/10.1016/j.cobeha.2017.05.009

Molnar, C. (2022). *Interpretable Machine Learning: A guide for making black box models explainable* (2nd ed.).

Montani, I., Honnibal, M., van Landeghem, S., Boyd, A., Peters, H [Henning], O'Leary McCann, P., Geovedi, J., O'Regan, J., Samsonov, M., Orosz, G., Kok, D. de, Altinok, D., Lind Kristiansen, S., Kannan, M., Bournhonesque, R., Miranda, L., Baumgartner, P., Edward, Explosion Bot, . . . Bozek, S. (2022). *explosion/spaCy: v3.4.4* [Computer software]. Zenodo.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-Factor Model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, *11*(2-3), 145–165. https://doi.org/10.1080/08959285.1998.9668029

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality & Social Psychology Bulletin*, *35*(12), 1661–1671. https://doi.org/10.1177/0146167209346309

Nave, G., Minxha, J., Greenburg, D., Kosinski, M., Stillwell, D. J., & Rentfrow, P. (2018). Musical preferences predict personality: Evidence from active listening and Facebook likes. *Psychological Science*, *29*(7). https://doi.org/10.17863/CAM.21182

Nye, C. D. (2022). Assessing Interests in the Twenty-First-Century Workforce: Building on a Century of Interest Measurement. *Annual Review of Organizational Psychology and Organizational Behavior*, *9*(1), 415–440. https://doi.org/10.1146/annurev-orgpsych-012420-083120

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *7*(4), 384–403. https://doi.org/10.1177/1745691612449021

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior*, *98*, 138–151. https://doi.org/10.1016/j.jvb.2016.11.002

Ostroff, C., & Zhan, Y. (2012). Person–environment fit in the selection process. In N. Schmitt (Ed.), *Oxford Handbook of Personnel Assessment and Selection* (pp. 252–273). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199732579.001.0001

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. https://doi.org/10.1037/pspp0000020

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.5555/1953048.2078195

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.* University of Texas at Austin.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

Peters, H [Heinrich], & Matz, S. (2023). Large Language Models can infer psychological dispositions of social media users. Advance online publication. https://doi.org/10.48550/arXiv.2309.08631

Roulin, N., & Levashina, J. (2019). LinkedIn as a new selection method: Psychometric properties and assessment approach. *Personnel Psychology*, *72*(2), 187–211. https://doi.org/10.1111/peps.12296

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, *26*(3), 565–606. https://doi.org/10.1177/014920630002600308

Sajjadiani, S., Daniels, M. A., & Huang, H.-C. (2022). The social process of coping with work-related stressors online: A machine learning and interpretive data science approach. *Personnel Psychology*, Article peps.12538. Advance online publication. https://doi.org/10.1111/peps.12538

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology*, *104*(10), 1207–1225. https://doi.org/10.1037/apl0000405

Salgado, J. F., & Fruyt, F. de. (2017). Personality in personnel selection. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *Handbooks in management. The Blackwell handbook of personnel selection* (pp. 174–198). Blackwell. https://doi.org/10.1002/9781405164221.ch8

Schaarschmidt, M., Walsh, G., & Ivens, S. (2021). Digital war for talent: How profile reputations on company rating platforms drive job seekers' application intentions. *Journal of Vocational Behavior*, *131*, 103644. https://doi.org/10.1016/j.jvb.2021.103644

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D. J., 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

Segalin, C., Perina, A., Cristani, M., & Vinciarelli, A. (2017). The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed. *IEEE Transactions on Affective Computing*, *8*(2), 268–285. https://doi.org/10.1109/TAFFC.2016.2516994

Spencer, D. A. (2018). Fear and hope in an age of mass automation: Debating the future of work. *New Technoloy, Work and Employment*, *33*(1), 1–12. https://doi.org/10.1111/ntwe.12105

Stachl, C., Boyd, R. L., Horstmann, K. T., Khambatta, P., Matz, S. C., & Harari, G. M. (2021). Computational personality assessment. *Personality Science*, *2*, Article e6115. https://doi.org/10.5964/ps.6115

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., Gosling, S. D., & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality*, *34*(5), 613–631. https://doi.org/10.1002/per.2257

Stoll, G., Rieger, S., Lüdtke, O., Nagengast, B., Trautwein, U., & Roberts, B. W. (2017). Vocational interests assessed at the end of high school predict life outcomes assessed 10 years later over and above IQ and Big Five personality traits. *Journal of Personality and Social Psychology*, *113*(1), 167–184. https://doi.org/10.1037/pspp0000117

Tay, L., Woo, S. E., Hickman, L., Booth, B. M., & D’Mello, S. (2022). A conceptual framework for investigating and mitigating Machine Learning Measurement Bias (MLMB) in psychological assessment. *Advances in Methods and Practices in Psychological Science*, *5*(1), 251524592110613. https://doi.org/10.1177/25152459211061337

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G.

(2023, February 27). *LLaMA: Open and Efficient Foundation Language Models*. https://arxiv.org/pdf/2302.13971

Twenge, J. M., Campbell, S. M., Hoffman, B. J., & Lance, C. E. (2010). Generational differences in work values: Leisure and extrinsic values increasing, social and intrinsic values decreasing. *Journal of Management*, *36*(5), 1117–1142. https://doi.org/10.1177/0149206309352246

van de Ven, N., Bogaert, A., Serlie, A., Brandt, M. J., & Denissen, J. J. (2017). Personality perception based on LinkedIn profiles. *Journal of Managerial Psychology*, *32*(6), 418–429. https://doi.org/10.1108/JMP-07-2016-0220

van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, *96*(6), 1167–1194. https://doi.org/10.1037/a0024343

Wegmeyer, L. J., & Speer, A. B. (2022). Understanding, detecting, and deterring faking on interest inventories. *International Journal of Selection and Assessment*, *30*(4), 562–578. https://doi.org/10.1111/ijsa.12398

Wilmot, M. P., & Ones, D. S. (2021). Occupational characteristics moderate personality–performance relations in major occupational groups. *Journal of Vocational Behavior*, *131*, 1–18. https://doi.org/10.1016/j.jvb.2021.103655

Yarbrough, J. R. (2018). Is cybervetting ethical? An overview of legal and ethical issues. *Journal of Ethical and Legal Issues*, *11*.

Youyou, W., Kosinski, M., & Stillwell, D. J. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112