Automatic Personality Prediction Based on Facial Features: Race, Gender, and Age Bias

A Study on the Impact of Demographic Covariates on a Facial Recognition System

Nele S. Keszler (s2061465)

Department of Research Methodology, Measurement and Data Analysis (OMD)

Faculty of Behavioural, Management and Social Sciences (BMS)

University of Twente

PSY-B-18: Bachelor Thesis

Prof. Dr. Ir. B. P. Veldkamp

Dr. P. Papadopoulos

L. Kinteh

June 25, 2021

Abstract

In modern times, new technological innovations change the way we live and work, how we perceive others and ourselves. Latest research in the personality assessment industry has considered machine learning for job screening processes. A specific, yet less investigated method for fully automated personality prediction is offered through facial recognition systems (FRS). In the context of machine learning algorithms developed for personality prediction, compliance with modern Diversity, Equity, and Inclusion standards of the industry must be considered. To do so, an unbiased nature of such machine learning algorithm needs to be aimed at. In this causal-comparative study, previously developed deep learning models for personality prediction based on facial features were tested for influence of race, gender, and age on the correctness of classifying high or low Extraversion and Conscientiousness using logistic regression analyses. Hereby, the research question 'To what extent does race, gender and age influence the prediction of Extraversion and Conscientiousness through an FRS?' is aimed at to be answered. Two stratified samples were used, with 75 and 85 participants respectively. None of the predictor variables showed a significant influence on correctness of prediction for either trait. This leads to the conclusion that the algorithm predicts Extraversion and Conscientiousness in an unbiased manner. For future research, it is advised to further validate the algorithm on new data and continuous score variables. To enable an adequate use of the FRS in its context, it should additionally be tested on more diverse samples and for other personality traits.

Contents

Introduction

The assessment of personality has been extensively addressed by psychologists for more than a century. Starting in World War I, standardized personality tests were created to find a suitable fit between (military) recruits and job requirements (Christiansen & Tett, 2013). From then on, personality assessment has developed into a $400 million industry with annual growth of 8 to 10 percent in the year 2010 (Paul, 2010). One decade later, the industry is now worth around $2 billion, with personality tests being used for the hiring process in numberless companies (Escalante, 2020). In connection with this, personality assessment offers numerous advantages for personnel selection, such as finding a good job fit, but also drawbacks, such as being time-consuming and costly (Bernstein, 2017).

Consequently, personality assessment undergoes constant refinement to tackle existing drawbacks, especially with the continuous development of new technologies. For job candidate screening, increasing numbers of contributions on the topic of machine learning have come forward. Special interest has been shown in the prediction of personality with the help of facial recognition systems (FRS), using, for example, convolutional neural networks (CNN) (Wei, Zhang, Zhang, & Wu, 2017; Ventura, Masip, & Lapedriza, 2017). These deep learning models are thereby extracting and analyzing features from image or video material of participants to assess specific traits (Escalera, Baró, Guyon, & Escalante 2018). Although this depicts automated personality prediction as a promising refined process, with possibly valuable opportunities for selecting employees (Liem et al., 2018), advantages, as well as drawbacks, still need to be discussed in the following.

To understand the context in which FRS would be implemented, we first need to have a look at trends and issues in modern society. The gender pay gap and the Black Lives Matter movement, for example, highlight existing societal problem areas of the 21st century that also show an impact on the hiring process (Ward & Heys, 2020). In Europe, a clear discrimination against, for example, job candidates of diverse ethnicities can still be found within organizations (Imdorf, 2017). Hence, at the workplace, managers and recruiters are currently addressing assessment procedures that might discriminate against specific minorities or marginalized groups of people. Such assessment procedures often include a bias or stereotypes that people involved in the evaluation process naturally hold, e.g., against specific ethnicities or genders (Hebl, Madera, & Botsford Morgan, 2019; Polli, 2019). Accordingly, fully programmed assessment methods using AI could bring solutions to these challenges, as predictive algorithms are not influenced by human bias (Polli, 2019). Here, a clear advantage or chance for improvement is established by automated assessment.

While automated assessment methods offer great advantages in theory, the application of AI in practice has shown some risks. To depict those risks more vividly, the following examples can be given: To begin with, as reported by Crawford (2016) in the New York Times, various image recognition software show racial biases. One example that is given in the article was camera software of different manufacturers that were unable to recognize dark faces or misinterpreted pictures of Asians as blinking. As a prominent reason for these problems, it is stated that predictive algorithms are solely dependent on the data they are trained on (Crawford, 2016). Hence, data showing bias in the training process will consequently result in a biased prediction. Therefore, using an FRS with such flaws for assessing job candidates could result in tremendous negative effects of discrimination against specific skin colors or ethnicities.

However, not only cases of racial discrimination must be considered but also equitable employment in terms of gender when using AI. At least since the Amazon.com Inc case of 2015, reported by Reuters News (Dastin, 2018), the importance of algorithms free of gender bias has become clear. After experimenting on an AI system to improve hiring processes in 2014, the company realized one year later that the system would be favoring male over female applicants for technical jobs and was therefore not gender neutral. This was because the algorithm was trained on résumés of prior job applicants in the male-dominated tech industry, in which a clear gender gap was observable.

The given examples highlight the relevance of ethical and technical considerations when adopting AI for the hiring process. To summarize considerations around employee selection with regards to avoiding bias, a common term that is used in the discussion of promoting equal hiring chances and inclusive work environments is DEI. DEI stands for Diversity, Equity, and Inclusion (Marinaki, n.d.), hence, creating diverse teams with equal chances. Therefore, taking a closer look at possible threats to DEI that an algorithm for personnel selection might hold seems crucial and justified at this point. Thus, the aim of the current study was to determine the extent of possible biases in personality assessment through automated personality prediction using FRS. Consequently, the question for research arises to what extent race, gender, and age influence personality prediction through an FRS.

Theoretical Background

Traditional Personality Assessment

To begin with, the concept of personality needs to be reviewed. Personality determines how each of us behaves, feels, and thinks (Funder, 2004). As each person possesses a distinct, relatively stable set of traits or characteristics, the concept of personality can be used to capture one's individuality (Maltby, Day, & Macaskill, 2010). Assessing personality has therefore been a popular research object for psychologists in the past. Starting in the mid-20th-century until present days, various researchers have identified a robust model of five factors that determines personality (Digman, 1990). Evolved from early inventories such as the three dimensions of Psychoticism, Extraversion, and Neuroticism, also known as PEN, (Eysenck, as cited in Digman, 1990) and the NEO PI-R (Costa, McCrae, & Kay, 1995), a popular general taxonomy of personality is the Five Factors Model (FFM), or Big Five, including the dimensions Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Goldberg, 1992; Saucier & Goldberg, 1996; Costa & McCrae, 1992).

In the Human Resource (HR) sector, this taxonomy has especially been of help for the hiring process. Rothmann and Coetzer (2003), for example, have found relations between employees' Big Five personality traits and their task performance, creativity, and management performance. Their results also show that with $R^2 = 0.28$, the five personality dimensions explain 28% of the variance in management skills, for example. In addition, Barrick and Mount (1991) found that the distinct dimensions of the Big Five show direct relations with different job performance criteria over diverse occupations. For example, in their study, Conscientiousness shows true score correlations ranging from $\rho = 0.20$ with the occupational group Professionals to $\rho = 0.23$ with the occupational group Sales. Conscientiousness is further revealed to correlate with all job performance criteria of their study, for example job proficiency ($\rho = 0.23$) and training proficiency ($\rho = 0.23$) (Barrick & Mount, 1991). Consequently, assessing the personality of job candidates, for example by testing their Big Five traits, has become a common tool for HR-professionals.

Personality Prediction Based on Facial Features

In a recent review on new technology in the field of personality assessment, Ihsan and Furnham (2018) have summarized promising inventions that aim at improving the assessment procedure, for example in terms of reduced costs and higher validity. One of the new technologies that is concerned in their review is automated personality testing, i.e., personality testing with the help of artificial intelligence (AI) in the form of machine learning algorithms. As described by Liem et al. (2018), the use of machine learning algorithms for job screening processes gained popularity because they show higher efficiency and scalability, compared to HR-professionals without computer assistance. In accordance with that, research in the field has gone in the direction of machine learning algorithms predicting personality based on facial features (Al Moubayed, Vazquez-Alvarez, McKay, & Vinciarelli, 2014; Biel, Teijeiro-Mosquera, & Gatica-Perez, 2012). Zhang (2013) describes face recognition as a popular and important biometric method, as it is easy and effective to use. In other words, FRS are shaping up well in the scientific society and seem to be theoretically approved by many.

Regarding the success in application of automated FRS, Rojas, Masip, Todorov, and Vitria (2011) have reported in their study that characteristics or traits of human personality can be accurately predicted from face information using different machine learning approaches. Accordingly, the ChaLearn Looking at People 2016 Challenge on First Impressions (Ponce-López et al., 2016) made substantial contributions to the field. In their challenge, different teams successfully approached automated personality prediction through video and audio material using deep learning models. Based on accuracy, the best performing models gave valuable insights, nonetheless, some methods performed better on specific traits than others. Overall, their results show that the traits of Extraversion and Conscientiousness could best be predicted, with explained variances higher than 50% ($R^2 = 0.51$ and $R^2 = 0.54$ respectively) and accuracy based on classification assessed through area under the receiver operating characteristic curve (ROC-AUC) of 0.87 for Conscientiousness and 0.83 for Extraversion. This is in line with further research, in which models performed best on Extraversion with $R^2 = 0.17$ (Teijeiro-Mosquera, Biel, Alba-Castro, & Gatica-Perez, 2014) and predicted Conscientiousness, with a significant Pearson correlation between observed and predicted values of $r = 0.360$ for men and $r = 0.335$ for women (Kachur, Davydov, Shutilov, & Novokshonov, 2020). The study of Junior et al. (2019), in which the different work of several researchers was analyzed, shows that Extraversion and Conscientiousness are overall the best recognized traits in personality prediction. Nevertheless, Ihsan and Furnham (2018) argue that ongoing research is vital in this novel field of automated personality testing as gathering more data would help to improve the technology and to predict personality more accurately. This raises questions about what confounding factors make automated personality prediction less accurate or which covariates are influencing the prediction of personality using FRS.

Possible Biases in FRS

Concerning possible biases in personality prediction based on facial features, studies about gender influence show differences in their results thus far. Various studies reveal that, for example, specific traits can more reliably be predicted from female faces compared to male faces (Kachur et al., 2020; Qin, Gao, Xu, & Hu, 2016). In the study by Hu et al. (2017), however, personality could be better recognized from male than from female faces. As their dataset consisted of only Asian participants, they point out that these findings are different from studies conducted on European subjects. These findings can be substantiated by the study of Escalante et al. (2018) who further analyzed the results of the ChaLearn Looking at People 2016 Challenge on First Impressions (Ponce-López et al., 2016). Their results reveal an existing bias in both gender and ethnicity compared to Big Five personality tests. To be more specific, a positive bias for females on the traits and job interview invitations and an overall positive bias for Caucasians compared to a negative bias towards African Americans is identified. Thus, not only the variable gender seems to be influential on personality prediction based on faces, but also the racial background of participants should be taken into account.

With regards to specific personality traits, the study of Zhang et al. (2017) on automated trait prediction using pictures of exclusively East-Asian participants revealed that a trait belonging to Conscientiousness can be accurately and reliably predicted by a deep learning model as opposed to Extraversion. Additionally, Hu et al. (2017) found among Han Chinese participants that Conscientiousness could be associated with specific patterns in male faces and Extraversion with patterns in female faces. From those findings, it can be concluded that participants' race might be an influencing factor when assessing Extraversion and Conscientiousness from faces. Respectively, Escalera et al. (2018) suggest that racial biases stemming from computational personality analysis through videos are a prominent problem for which effective solutions must be researched.

Besides gender and race, Abdurrahim, Samad, and Huddin (2018) summarized in their review that age constitutes a third demographic covariate with effects on FRS, as in most studies, older faces are more easily recognized than younger faces. Furthermore, Raja et al. (2020) highlighted that especially ethnicity, gender, and age are three demographic covariates that influence the performance of algorithms programmed for facial recognition. Therefore, considering the importance of DEI in the workplace today, it is suggested that in the ongoing research and development, FRS algorithms need to be tested – and become more accurate – with regards to especially these three possible covariates.

In summary, it can be stated that novel ways of personality assessment lay the ground for research on the further development and improvement of such. As previously stated, the goal of this study was to examine to what extent race, gender, and age influence personality prediction through an FRS. To refine that goal based on the above-reviewed literature, it is hypothesized that race, gender, and age significantly influence the automated prediction of Extraversion as well as Conscientiousness measured with an FRS. In accordance, the following research question was formulated: 'To what extent does race, gender, and age influence the prediction of Extraversion and Conscientiousness through an FRS?'. To answer this research question, a causal-comparative, or ex post facto, study was conducted.

Method

Participants

For the current study, ethical approval was obtained from the Ethics Committee of the University of Twente. Two stratified samples were used, one for each personality trait that was investigated. In the beginning, the following information was collected: ethnicity, gender, and age whereby the participants indicated to be either male (0) or female (1). Participants' ethnicities ranged from Irish, Polish, Malawian, Ghanaian, Zambian, Australian, Canadian, Zimbabwean, Nigerian, US American, South African, to British, which were recoded into White/Caucasian (0) or Black/African American (1) based on skin color, to distinguish participants for analyses on race. The first dataset was used for statistical analyses on the trait of Extraversion and consisted of 75 participants. From those, 38 participants were female, and 37 participants were male (M = 0.51, SD = 0.50). Furthermore, 37 participants had dark skin color, i.e., were Black/African American and 38 participants had light skin color, i.e., were White/Caucasian (M = 0.49, SD = 0.50). Besides that, the age ranged from 19 to 49 years (M = 31.75, SD = 7.58) with the greatest proportion of 8% being 29 years of age.

The second dataset was used for statistical analyses on the trait of Conscientiousness and consisted of 85 participants. From those, 42 participants were female, and 43 participants were male (M = 0.49, SD = 0.50). Moreover, 38 had dark skin color, i.e., were Black/African American and 47 participants had light skin color, i.e., were White/Caucasian (M = 0.45, SD = 0.50). Besides that, the age ranged from 18 to 49 years (M = 32.82, SD = 7.43) with the greatest proportion of 8.2% being 29 years of age.

Materials

The instruments used for the current study were the following: (1), a short scale assessing the Big Five, called the Big-Five-Inventory-10 (BFI-10) by Rammstedt, Kemper, Klein, Beierlein, and Kovaleva (2012), (2), three gamified neuro-assessments of cognitive ability and personality from the Dutch company Zyvo (https://www.zyvo.nl/en), (3), one job-related interview question. During the completion of gamified neuro-assessments and the interview question, videos of participants' faces were recorded without audio. On the basis of this video material, the deep learning models for personality prediction based on facial features were programmed by Röber (2021) and subsequently used as the main instrument of the current study.

BFI-10

A validated personality test was used to measure true scores on Extraversion and Conscientiousness of the participants. The test was a ten-item short scale assessing the Big Five, called BFI-10 (Rammstedt et al., 2012). The BFI-10 consisted of two items for the assessment of each of the five personality traits Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. It shows questionable reliability with retest-coefficients of 0.49 to 0.59 per scale on a sample representing the general population but was tested for its content validity, factor validity and construct validity, showing adequate outcomes.

Gamified Neuro-Assessments, Interview Question and Video Recordings

The gamified neuro-assessments provided by Zyvo tested participants' personality and cognitive ability. To begin with, 'Balloon' is a game designed to assess the personality facets of risk-taking ability, loss aversion and need for excitement. Next, the game 'Code compare' aims at assessing participants' mental ability of perceptual speed and their accuracy and lastly, 'The Switch' tests the preference and ability of participants to switch between tasks. During the process of completing these games as well as answering the subsequent interview question, participants' faces were recorded through their webcam.

Facial Recognition System

Subsequently, algorithms were programmed to predict Extraversion and Conscientiousness from facial features, as described by Röber (2021). The algorithms were trained based on video material of the participants' faces. First, in both cases, machine learning methods were used for face detection for each frame of a video as well as facial landmark detection. In the next step, the faces were aligned on a horizontal axis, by rotating and scaling them uniformly for all frames. On the basis of (x,y)-coordinates of facial landmarks, identified features, such as (x,y)-landmarks describing the eyes or variation in pupil location, were extracted. Finally, model exploration was performed using different of these generated features as input (Röber, 2021). To avoid bias during the explorative training process, the training data, comprising 80% of the available data used in this study, was split from the testing data, or hold-out set, comprising the remaining 20% of the available data which was used to assess model performance. Important to note is that during the explorative training process, the algorithm was adjusted and improved in various iterative cycles by Röber (2021), until the best performing models, used in the current study, arrived at 87% and 94% accuracy on the hold-out test set for Extraversion and Conscientiousness respectively.

Both final deep learning models, one for Extraversion and one for Conscientiousness, utilized (x,y)-coordinates of landmarks describing the eyes only as input features and were constituted of a CNN with one convolutional layer followed by two dense layers each. For Extraversion, the best performing model was based on video material recorded during the interview question and for Conscientiousness, the best performing model was based on video material recorded during the game ‘The Switch’. Features used as input for both models stem from videos of around 31 seconds with 15 frames per second. The output of both models was a predicted probability for each trait, where predicted probability > 0.5 is classified as scoring ‘high’ on the trait (classified as ‘1’) and predicted probability < 0.5 is classified as scoring ‘low’ on the trait (classified as ‘0’). For this study, the classification into ‘high’ and ‘low’ scoring on each trait was decided for as it gives most information needed for its application. As summarized by Stidham, Summers, and Shuffler (2018) scoring ‘high’ in Extraversion could be interpreted as being outgoing and warm whereas scoring ‘low’ would mean to be quieter and more reserved, and scoring ‘high’ in Conscientiousness could be interpreted as being hardworking, dependable, and organized whereas scoring ‘low’ would mean being more careless and disorganized. Hence, as the FRS was developed primarily for the selection of employees, a first investigation on the more straightforward ‘high’ or ‘low’ classification seemed most appropriate, because it could separate job candidates into better or worse fitting for the position or organization.

Design

As described above, an FRS was operationalized by using methods for face detection and facial landmark detection, on which basis deep learning models for Extraversion and Conscientiousness prediction were programmed and trained by Röber (2021), prior to the current study. For the present ex post facto study, a between-groups design was then applied for these personality prediction models. The independent variables were gender, age, and race of participants. The dichotomous dependent variables were correctly classified Extraversion scores through the personality prediction model and the correctly classified Conscientiousness scores through the personality prediction model with '1' indicating the classification was correct and '0' indicating the classification was incorrect.

Procedure

The data collection was performed by the company Zyvo in the period between December 2020 and March 2021 through a global online research platform. When participants signed up on the platform voluntarily and chose to participate in the study, first the aim of the study was explained. Afterwards, the participants were asked to give informed consent (e.g., specifically about video recording through their webcam; no audio was recorded) and to indicate their age, ethnicity, and gender to enable the creation of balanced stratified samples post study. Next, the participants were redirected to start with completing the BFI-10. Subsequently, their face was recorded through their webcam while playing Zyvo's gamified neuro-assessments 'Balloon', 'Code compare' and 'The Switch' and answering one job-related interview question. The study took approximately 30 minutes in total per participant and was instructed and completed in English language.

Data Analyses

The deep learning models were developed using Python (Version 3.8.3) to derive predictions of Extraversion and Conscientiousness and afterwards downloaded and analyzed using IBM SPSS Statistics (Version 25) predictive analytic software for the ex post facto study.

Trait Prediction

First, the distribution of BFI-10 scores on Extraversion and Conscientiousness were investigated. True scores were then recoded into 'high' (1) when being above the mean and 'low' (0) when being below the mean, based on reference values by Rammstedt et al. (2012).

Next, the FRS was run in Google Colab based on the input file of all (x,y)-coordinates of extracted landmarks and the related personality prediction models (see Materials). This resulted in a predicted probability of scoring high on Extraversion or Conscientiousness respectively to a specific degree. Next, these predicted probabilities were included as a continuous variable in the associated original data sets. Then, the predicted probabilities were classified into a dichotomous variable as the final predicted score, with predicted probability > 0.5 being classified as '1' and predicted probability < 0.5 being classified as '0'. This step classified the predicted degree of the trait into a being predicted as either 'high' (1) or 'low' (0) in either trait. This classification was important for the sake of comparison between the predicted degree of traits to the before classified true scores in the last step. As a result, the dichotomous variable of correct classification was added to the dataset, showing whether the classification coming from the FRS was correct (1) or incorrect (0), i.e., whether the final classified predicted score equals the classified true score. This process was executed for both traits independently on their particular model and dataset. Lastly, the final datasets were downloaded and further analyzed in SPSS.

Generalized Linear Model (GLM)

In the next step, a GLM was utilized to investigate possible influences of the categorical independent variables race and gender and the continuous independent variable age on the dichotomous classification correctness of the FRS. To do so, two assumptions about the data needed to be checked first. If both assumptions are sufficiently fulfilled, a GLM could be constructed. The first assumption that was checked was the absence of multicollinearity by inspecting inter-correlations between the three predictor variables race, gender, and age. Multicollinearity was further checked by calculating the variance inflation factor (VIF) as suggested by Bhandari (2020). The next assumption that was checked was the absence of outliers by plotting the standardized residuals of the dependent variable against its standardized predicted value with a cut-off score of 3.3.

After confirming the assumptions, a GLM could be constructed. As the dependent variable of correctly classified traits was of dichotomous nature, a binary logistic regression was performed. Thus, a logistic regression model was constructed to investigate the influence of the independent variables on the probability of classifying traits correctly. The logit of the probability of correctly classifying a trait with the FRS is predicted, according to Van den Berg (2019), by the following linear model:

$$Logit_i = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

$$y \sim Bern(p_i)$$

where	$logit_i$	is the logarithm of the odds to correctly classify a specific trait
	y	is the observed success of the FRS
	$Bern(p_i)$	is the Bernoulli distribution with probability p of correctly classifying a specific trait with the FRS

Corresponding to the probability of correctly classifying either trait with the FRS and with independent variables race, gender, and age, this can be defined by the following formula:

$$ln \frac{p_i}{1 - p_i} = b_0 + b_1 race + b_2 gender + b_3 age$$

$$y \sim Bern(p_i)$$

It was hypothesized that all three independent variables were significant predictors for the probability of correct classifications for Extraversion and Conscientiousness, and hence, that the regression models outperform the null models. Additionally, The Hosmer-Lemeshow test was performed to make statements about the models' goodness-of-fit. Furthermore, Nagelkerkes R^2 was calculated to determine the proportion of variance in the categorical dependent variable that is explained by the model including race, gender, and age as predictor variables. To make statements about the results' significance, the critical value is chosen at $\alpha = 0.05$ with an according confidence interval of 95%.

Results

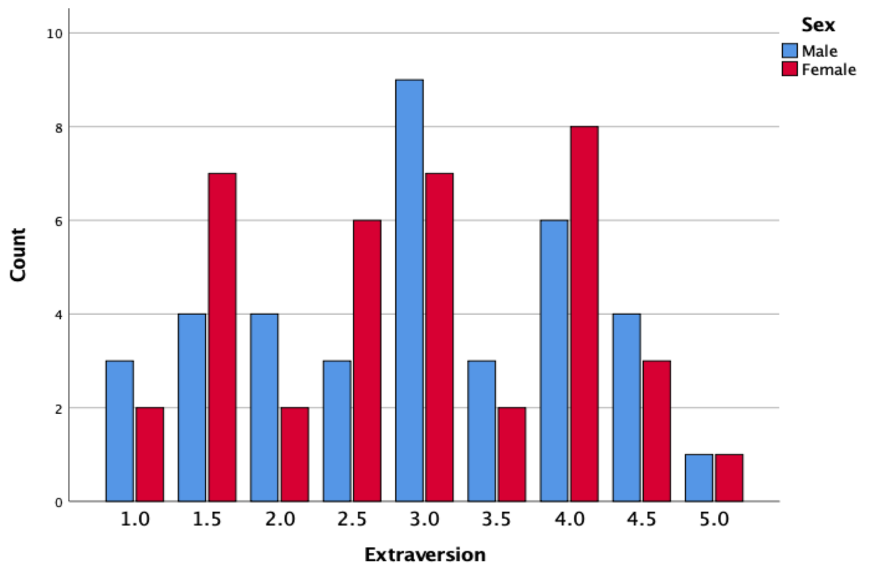
BFI-10 Scores

To begin with, descriptive statistics on the scores of participants' Extraversion and Conscientiousness assessed with the BFI-10 were investigated. The overall distribution of Extraversion scores for the sample seems normally distributed (Figure 1). Furthermore, participants' Extraversion scores can be compared to the norm groups established by Rammstedt et al. (2012). Their reference values based on a sample of 1134 participants for

Extraversion are the following: males aged 18 to 35 show Extraversion scores with *M* = 3.66 and *SD* = 0.86, whereas females of the same age group show Extraversion scores with *M* = 3.81 and *SD* = 0.90. In the current sample, 26 males aged 18 to 35 show Extraversion scores with *M* = 2.79 and *SD* = 1.16 and 29 females aged 18 to 35 show Extraversion scores with *M* = 2.85 and *SD* = 1.04. This distribution can be seen in Figure 2. Besides that, the reference values for males aged 36 to 65 show Extraversion scores with *M* = 3.45 and *SD* = 0.91, whereas females of the same age group show Extraversion scores with *M* = 3.53 and *SD* = 1.01. In the current sample, 11 males aged 36 to 49 show Extraversion scores with *M* = 3.32 and *SD* = 0.98, whereas the remaining 9 females of the same age group show Extraversion scores with *M* = 3.06 and *SD* = 1.38. This distribution can be seen in Figure 3. Concludingly, in the current sample, young males and females as well as middle-aged females have on average scored lower on Extraversion than the norm group, whereas middle-aged males have scored more similarly to their reference values, which might be because reference values were constituted not based on an international but German population.

Figure 1

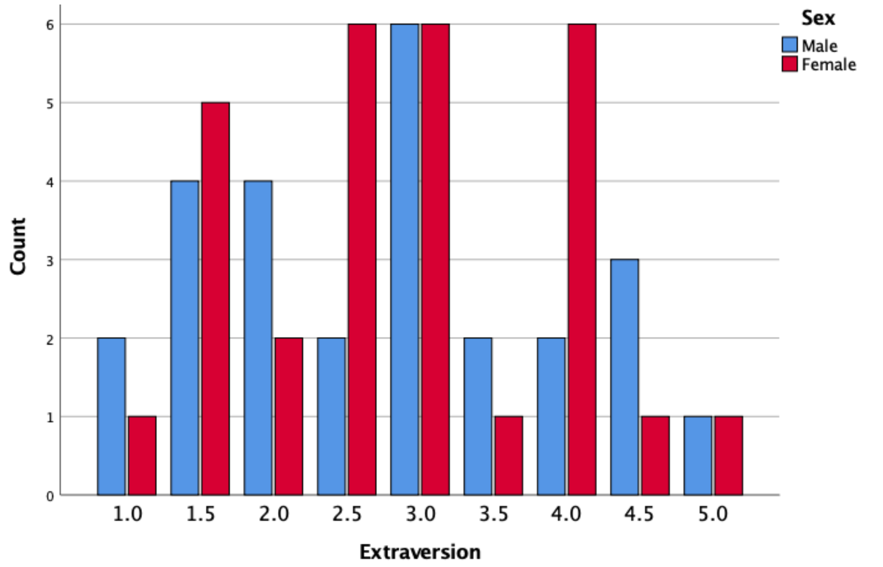
Overall Extraversion Scores



Note. Extraversion scores are shown for male and female participants for a sample of n = 75 and were administered with the BFI-10 by Rammstedt et al. (2012).

Figure 2

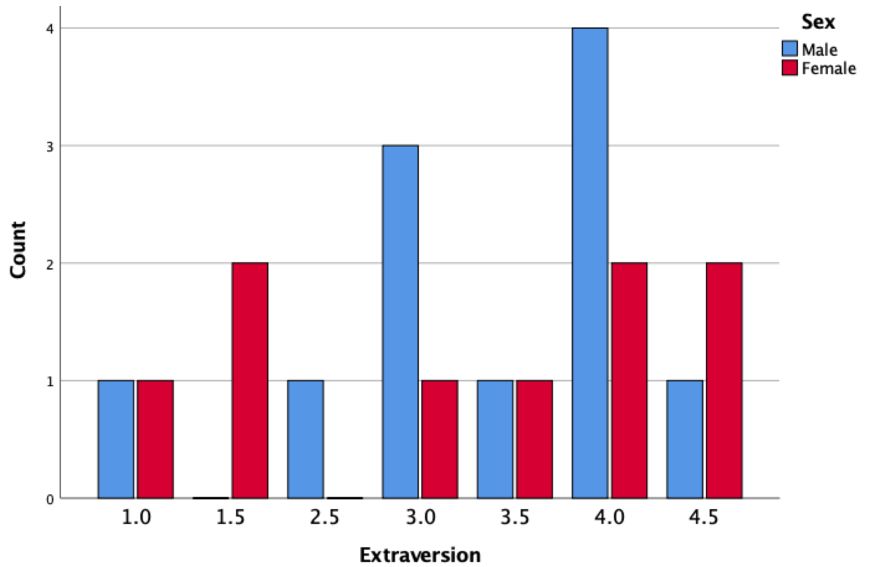
Extraversion Scores of Young Participants



Note. Extraversion scores are shown for male and female participants aged 18 to 35 for a sample of n = 55 and were administered with the BFI-10 by Rammstedt et al. (2012).

Figure 3

Extraversion Scores of Middle-Aged Participants

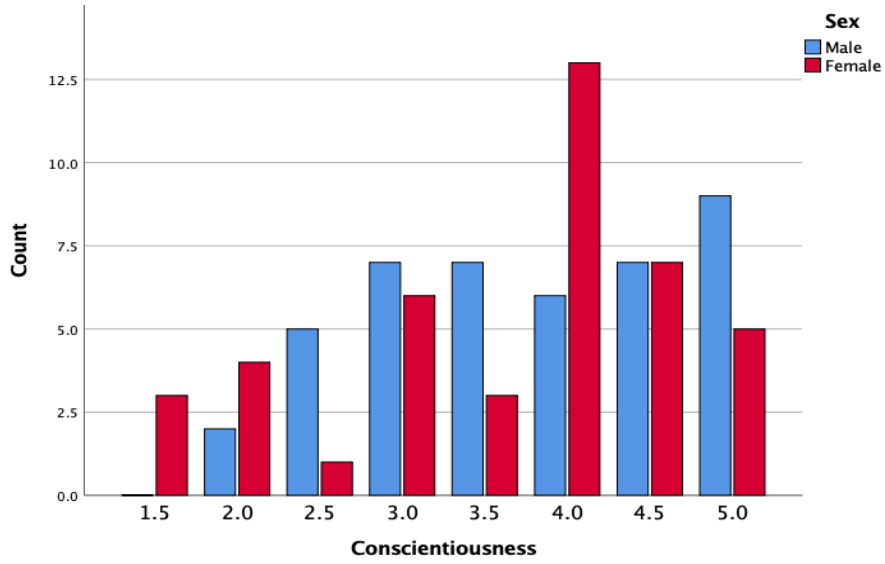


Note. Extraversion scores are shown for male and female participants aged 36 to 49 for a sample of n = 20 and were administered with the BFI-10 by Rammstedt et al. (2012).

The overall distribution of Conscientiousness scores for the sample seems slightly left skewed (Figure 4). Furthermore, participants' Conscientiousness scores can again be compared to the norm groups established by Rammstedt et al. (2012). Their reference values based on a sample of 1134 participants for Conscientiousness are the following: males aged 18 to 35 show Conscientiousness scores with $M = 3.78$ and $SD = 0.75$, whereas females of the same age group show Conscientiousness scores with $M = 4.02$ and $SD = 0.97$. In the current sample, 27 males aged 18 to 35 show Conscientiousness scores with $M = 3.59$ and $SD = 0.84$ and 29 females aged 18 to 35 show Conscientiousness scores with $M = 3.60$ and $SD = 0.93$. This distribution can be seen in Figure 5. Besides that, the reference values for males aged 36 to 65 show Conscientiousness scores with $M = 4.02$ and $SD = 0.84$, whereas females of the same age group show Conscientiousness scores with $M = 4.28$ and $SD = 0.70$. In the current sample, 16 males aged 36 to 49 show Conscientiousness scores with $M = 4.09$ and $SD = 1.02$, whereas the remaining 13 females of the same age group show Conscientiousness scores with $M = 3.65$ and $SD = 1.30$. This distribution can be seen in Figure 6. Concludingly, in the current sample, young and middle-aged males scored similarly to their reference values and young as well as middle-aged females have on average scored lower on Conscientiousness than the norm group, which, once again, might be because reference values were constituted not based on an international but German population.

Figure 4

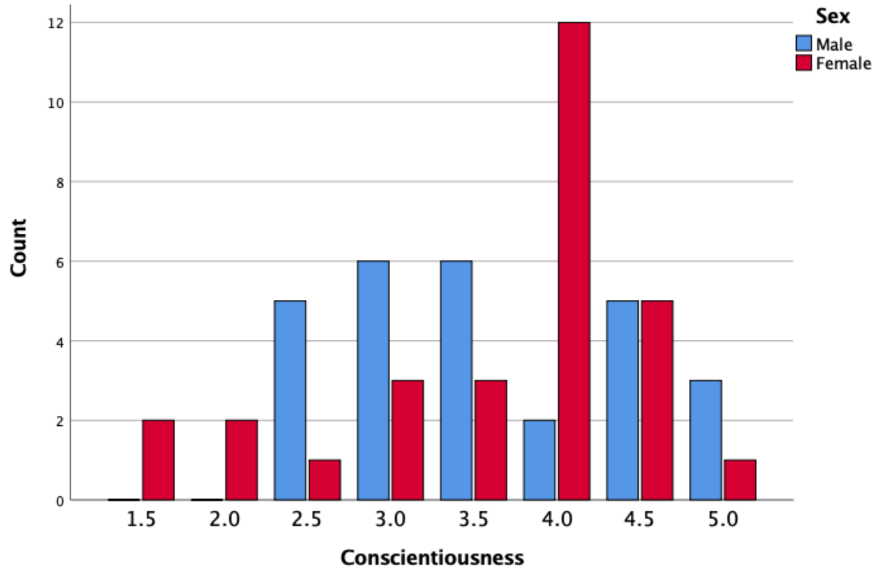
Overall Conscientiousness Scores



Note. Conscientiousness scores are shown for male and female participants for a sample of n = 85 and were administered with the BFI-10 by Rammstedt et al. (2012).

Figure 5

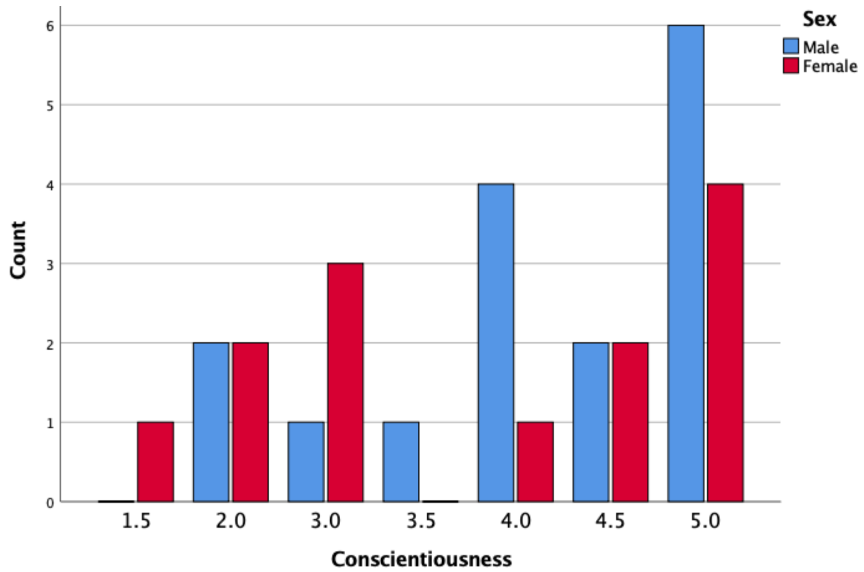
Conscientiousness Scores of Young Participants



Note. Conscientiousness scores are shown for male and female participants aged 18 to 35 for a sample of n = 56 and were administered with the BFI-10 by Rammstedt et al. (2012).

Figure 6

Conscientiousness Scores of Middle-Aged Participants



Note. Conscientiousness scores are shown for male and female participants aged 36 to 49 for a sample of n = 29 and were administered with the BFI-10 by Rammstedt et al. (2012).

Logistic Regression Analyses

To start with the model assumptions, bivariate correlations between the three predictor variables showed no significant relationship. In addition, the VIF was below 10 in both samples, showing that no multicollinearity was present. Lastly, no considerable outliers could be found in the data. Accordingly, a logistic regression analysis to investigate the extent of race, gender, and age influence on the correctness of Extraversion prediction with an FRS was performed first. The outcome of interest was correctly classified Extraversion class, and the predictor variables were race, gender, and age. The Hosmer-Lemeshow goodness-of-fit was significant ($p < 0.05$) indicating the model including the three independent variables was not a good fit and did not outperform the null model. Additionally, the Nagelkerke $R^2 = 0.047$. None of the predictor variables in the logistic regression analysis were found to significantly contribute to the model (Table 1). The model had an unstandardized Beta weight for the constant of $B = 2.937$, $SE = 2.333$, $Wald\ \chi^2 = 1.584$, $p > 0.05$. For the second logistic regression analysis, the dependent variable was correctly classified Conscientiousness class with predictor variables race, gender, and age. The Hosmer-Lemeshow goodness-of-fit was significant ($p < 0.05$) indicating the model including the three independent variables was not a good fit and did not outperform the null model. Additionally, the Nagelkerke $R^2 = 0.043$. None of the predictor variables in the logistic regression analysis were found to significantly contribute to the model (Table 1). The model had an unstandardized Beta weight for the constant of $B = 1.979$, $SE = 1.526$, $Wald\ \chi^2 = 1.683$, $p > 0.05$.

Table 1

Logistic regression analysis predicting correct classification of predicted personality traits

Predictor	B	*SE*	*Wald* χ^2	Exp(B)	95% CI for Exp(B)	
					LL	*UL*
			Extraversion			
Age	-.016	.065	.057	.985	.887	1.118
Gender	1.175	1.180	.992	3.240	.321	32.746
Race	-.013	.000	.000	.988	.128	7.629
			Conscientiousness			
Age	-.007	.041	.033	.933	.916	1.075
Gender	.399	.588	.461	1.491	.471	4.721
Race	-.745	.598	1.552	.475	.147	1.532

Note. SE = standard error; CI = confidence interval; *LL* = lower limit *UL* = upper limit.
* $p < 0.05$

Discussion

The current study examined the extent of race, gender, and age influence on the prediction correctness of a priorly developed FRS. The above-illustrated findings on the utilized FRS offer new, potentially valuable insights into the field of automated personality prediction. All three demographic variables were revealed as unrelated to the correct classification of the traits Extraversion and Conscientiousness. Therefore, the hypothesis that all three variables would influence the prediction outcomes on both traits, stemming from a prior thorough literature review, needs to be rejected. To answer the research question, it can be stated that age, gender, and race do not influence the prediction of Extraversion and Conscientiousness with the here utilized FRS. Hence, the FRS seems to show no bias in form of age, gender, and race differences when statistically investigating correct classification of both traits based on the available data. Consequently, in its context of use, the correct prediction of scoring ‘high’ or ‘low’ on Extraversion and Conscientiousness by the FRS could be interpreted as in line with DEI standards based on these analyses. The findings of the current study, however, are contradictory to the literature on FRS bias of recent years. As described by

Leslie (2020) numerous FRS using machine learning approaches that were tested reveal bias based on demographics for various reasons besides systemic ones, such as "discrimination arising from unbalanced data sampling, collection and labelling practices, skewed datasets and biased data pre-processing and modelling approaches" (p. 15). As it is challenging to counteract all possible and common obstacles involved in developing machine learning models for facial recognition tasks, the current findings of an unbiased FRS are depicted as surprising and should be treated with caution. This gives reason to discuss implications and limitations of the study in the following.

Limitations and Recommendations

Due to the unexpected findings in connection with literature, clear limitations of the study design need to be further discussed. To begin with, one reason for the good performance of the models in the current study might be that 80% of the available data were used for training of the algorithms itself, as described above. The splitting of the data set, as suggested by Vabalas, Gowen, Poliakoff, and Casson (2019), was sufficient for initial testing of the algorithms during the training process. For the ex post facto study, however, it was not possible to collect additional data due to high costs and limited time, so it had to be dealt with the available data consisting of training and hold-out sets, to reach total sample sizes of $n = 75$ and $n = 85$ which would be large enough to ensure statistical power of the analyses. As 80% of the data used in this study was therefore identical with the training sets, the "unbiased" nature of the algorithms might not be surprising after all, since the correct classifications might partially be stemming from the training itself. A clear statement about influence of age, gender, and race on the FRS can thus not be made without further testing.

To prevent these possible issues in future research, collecting a new set of data for validation might be the best solution. This would also be in line with the suggestion of Crawford (2016), that bias which might be present in training data needs to be detected and sorted out, so that the models are no longer affected by it. To elaborate on that, by using a validation dataset after the training and initial testing of the models, overfitting could additionally be prevented or limited, so that the models do not learn from noise present in the training data (Brownlee, 2016). Further validation of the two utilized machine learning models through different balanced datasets is therefore recommended.

Furthermore, an influencing factor on the analyses' outcomes could have been the age range that was investigated. Literature suggests that FRS are prone to recognition differences when comparing old to young participants, which in other studies is defined by groups of teens to 30-year-olds against 40 to 60-year-olds or above (Givens, Beveridge, Draper, & Bolme, 2003; Givens, Beveridge, Draper, Grother, & Phillips, 2004). In this study, however, the age range showed a maximum at 49 years in both cases. On that basis, an ultimate claim about age bias-free prediction can hardly be made. The same argument is persistent for the ethnical groups that were investigated, being limited to Caucasian and African American/Black participants. For the validation of an age and racial bias-free FRS, this scope should be broadened to include more ethnicities, e.g., Asians as well as additional participants of age 50 and above.

Moreover, a limitation in the current study was given by the benchmarking instrument BFI-10. As mentioned before, the assessment that was utilized to make claims about the correct or incorrect prediction shows questionable reliability when tested on a sample representing the general population (Rammstedt et al. 2012). Therefore, it is recommended to use an instrument with high reliability and validity to increase the statistical strength of the basis for future ex post facto studies.

Lastly, in the current study, dichotomous classifications of a ‘correct’ or ‘incorrect’ prediction of ‘high’ or ‘low’ Extraversion and Conscientiousness with the FRS were made. However, to gain a broader insight into the extent of influence that covariates such as race, gender, and age have, continuous variables could be used. In that case, raw scores predicted with an FRS could first be compared to the raw scores of a validated instrument and afterwards tested for covariate bias. Furthermore, the analyses should be applied to all Big Five traits instead of limiting it to two. Overall, this might result in more specific insights into the FRS functioning.

In conclusion, the scientific community is advised to conduct additional research with consideration of the above-mentioned statistical adjustments on the scope and methods used. Furthermore, for the FRS’ operationalization in its context of use, attention must be paid to the nature of variables that are object of investigation and the instruments that are used for the validation, e.g., the reliability and validity of personality assessments that function as benchmark values. Ongoing research on this topic would be valuable to fill scientific gaps as well as for the optimization of a product used in line with its industry’s standards.

Ethical Implications

Implications and considerations from the current study, especially of ethical nature, should be reflected on to arrive at a thorough conclusion. For any facial recognition algorithm that is being programmed, the scientific community needs to be aware of responsibilities that this entails. Not only an unintended bias in the algorithms, which was investigated in the current study, should be highlighted as problematic but also the utilization of such algorithms for discriminatory reasons. As described by Van Noorden (2020), various FRS have been developed, peer-reviewed and published quite recently that were used for systemic racism against specific minorities from a political background. Consequently, this research topic has become subject to a broader human rights debate impacting the scientific community directly. For the current study this implies that although the development of algorithms used to distinguish between certain groups might be stemming from good intention, i.e., to advance procedures for employee selection as it is the case here, users and researchers in the field must be aware of the ways FRS could be utilized for unethical reasons and be sensible enough to prevent it.

Besides that, for prospective research on the presented machine learning models, the data collected to train and test FRS algorithms must be in line with laws governing data protection and data security. In recent events, where this has not been the case, multiple data sets of face images were distributed and used without informed consent (Van Noorden, 2020). This led to the fact that algorithms were trained on peoples' faces without them giving permission on it, besides the fact that released data sets initially collected for non-commercial studies could be used by companies for various purposes. As rules for the use of face images are not clear for every country and in every case (Van Noorden, 2020), it can only be advice to insure against ethical issues in this context based on Ethics Codes for research and from a legal standpoint.

As a last ethical implication for an FRS utilization in practice, the topic of interpretable and explainable machine learning, as discussed by Piano (2020), needs to be broached. Since the algorithms of the current study, comprising CNN, are of explainable nature, it would be too complex for the end-user to comprehend where the results it gives are specifically stemming from (O'Sullivan, 2020). However, it is clear by now that the FRS results should be transparent and somewhat interpretable by the person being dependent on its assessment, especially to ensure decisions that are in line with DEI standards. This means, for a best possible support of HR professionals through an FRS in its context of use, the explainable machine learning algorithms of the current study might need improvement in their interpretability. More specifically, this would mean that if a recruiter perceived a job candidate as outgoing and warm, but the algorithm assesses them as 'low' in Extraversion, the recruiter should be able to understand where this result is coming from, whether it was stemming from a precise part of the candidate's eyes or rather their gender, to make an informed decision. The ethical issue of transparency that is described by Piano (2020) is therefore crucial for the application of FRS for personality assessment in the future.

The given points to ponder highlight the ongoing ethical discussion surrounding FRS. In closing, it can be stated that facial recognition algorithms, especially with respect to personality assessment, offer plenty of opportunities for future research. Nonetheless, a main responsibility involved in this progress are considerations of methodological as well as ethical nature.

References

Abdurrahim, S. H., Samad, S. A., & Huddin, A. B. (2018). Review on the effects of age, gender, and race demographics on automatic face recognition. *The Visual Computer*, *34*(11), 1617-1630. https://doi.org/10.1007/s00371-017-1428-z

Al Moubayed, N., Vazquez-Alvarez, Y., McKay, A., & Vinciarelli, A. (2014). Face based automatic personality perception. *Proceedings of the 22nd ACM international conference on Multimedia*, 1153-1156. https://doi.org/10.1145/2647868.2655014

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

Bernstein, R. (2017, February 13). Pros and Cons of Personality Tests for Employment [Blog post]. Retrieved from https://online.concordia.edu/business-news/pros-and-cons-of-personality-tests/

Biel, J. I., Teijeiro-Mosquera, L., & Gatica-Perez, D. (2012). Facetube: predicting personality from facial expressions of emotion in online conversational video. *Proceedings of the 14th ACM international conference on Multimodal interaction*, 53-56. https://doi.org/10.1145/2388676.2388689

Brownlee, J. (2016, March 21). Overfitting and Underfitting With Machine Learning Algorithms [Blog post]. Retrieved from https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/

Christiansen, N., & Tett, R. (2013). *Handbook of personality at work*. Retrieved from https://books.google.de/books?isbn=113405579X

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), 5-13. https://doi.org/10.1037/1040-3590.4.1.5

Costa, P. T., McCrae, R. R., & Kay, G. G. (1995). Persons, Places, and Personality: Career Assessment Using the Revised NEO Personality Inventory. *Journal of Career Assessment*, *3*(2), 123-139. https://doi.org/10.1177/106907279500300202

Crawford, K. (2016, June 25). Artificial Intelligence's White Guy Problem. *The New York Times*. Retrieved from https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

Dastin, J. (2018, October 11). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, *41*(1), 417-440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Escalante, A. (2020, June 3). *Three Key Takeaways From A New Book Taking On The $2 Billion Personality Testing Industry.* Retrieved from https://www.forbes.com/sites/alisonescalante/2020/06/03/this-new- book-takes-on-the-2-billion-personality-testing-industry-a-review/?sh=487ad94d7b10

Escalante, H. J., Kaya, H., Salah, A. A., Escalera, S., Gucluturk, Y., Guclu, U., ... & van Lier, R. (2018). Explaining first impressions: modeling, recognizing, and explaining apparent personality from videos. *Preprint submitted for publication.* Retrieved from https://arxiv.org/abs/1802.00745

Escalera, S., Baró, X., Guyon, I., & Escalante, H. J. (2018). Guest editorial: apparent personality analysis. *IEEE Transactions on Affective Computing*, *9*(3), 299-302. doi: 10.1109/TAFFC.2018.2864230

Funder, D. C. (2004). The study of the person. In A. Javsicas (Ed.), *The Personality puzzle* (pp. 3-16). New York, NY: W. W. Norton.

Givens, G., Beveridge, J. R., Draper, B. A., & Bolme, D. (2003). A Statistical Assessment of Subject Factors in the PCA Recognition of Human Faces. *Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition Workshop, 8,* 96-114. doi: 10.1109/CVPRW.2003.10088

Givens, G., Beveridge, J. R., Draper, B. A., Grother, P., & Phillips, P. J. (2004). How features of the human face affect recognition: a statistical comparison of three face recognition algorithms. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2,* 381-388. doi: 10.1109/CVPR.2004.1315189

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment*, *4*(1), 26-42. https://doi.org/10.1037/1040-3590.4.1.26

Hebl, M., Madera, J. M., & Botsford Morgan, W. (2019). Special Issue on Reducing Discrimination in the Workplace: An Introduction. *Personnel Assessment and Decisions*, *5*(2), i-iii. https://doi.org/10.25035/pad.2019.02.001

Hu, S., Xiong, J., Fu, P., Qiao, L., Tan, J., Jin, L., & Tang, K. (2017). Signatures of personality on dense 3D facial images. *Scientific reports*, *7*(1), 1-10. https://doi.org/10.1038/s41598-017-00071-5

Ihsan, Z., & Furnham, A. (2018). The new technologies in personality assessment: A review. *Consulting Psychology Journal: Practice and Research*, *70*(2), 147-166. https://doi.org/10.1037/cpb0000106

Imdorf, C. (2017). Understanding discrimination in hiring apprentices: how training companies use ethnicity to avoid organisational trouble. *Journal of vocational education & training*, *69*(3), 405-423. https://doi.org/10.1080/13636820.2016.1278397

Junior, J. C. S. J., Güçlütürk, Y., Pérez, M., Güçlü, U., Andujar, C., Baró, X., ... & Escalera, S. (2019). First impressions: A survey on vision-based apparent personality trait analysis. *IEEE Transactions on Affective Computing*. *Manuscript accepted for publication.* doi: 10.1109/TAFFC.2019.2930058

Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the Big Five personality traits using real-life static facial images. *Scientific reports*, *10*(1), 1-11. https://doi.org/10.1038/s41598-020-65358-6

Leslie, D. (2020). Understanding bias in facial recognition technologies: an explainer. *The Alan Turing Institute*. https://doi.org/10.5281/zenodo.4050457

Liem, C. C. S., Langer, M., Demetriou, A., Hiemstra, A. M. F., Wicaksana, A. S., Born, M. P., & König, C. J. (2018) Psychology Meets Machine Learning: Interdisciplinary Perspectives on Algorithmic Job Candidate Screening. In H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, & M. van Gerven (Eds.), *Explainable and Interpretable Models in Computer Vision and Machine Learning.* (pp. 197-253). Cham, Switzerland: Springer.

Maltby, J., Day, L., & Macaskill, A. (2010). *Personality, Individual Differences and Intelligence.* Retrieved from https://books.google.de/books?isbn=0273722905

Marinaki, A. (n.d.). *DEI in the workplace: Three experts discuss.* Retrieved 2021, April 2, from https://resources.workable.com/stories-and-insights/dei-in-the-workplace-three-experts-discuss

O'Sullivan, C. (2020, September 17). Interpretable vs Explainable Machine Learning [Blog post]. Retrieved from https://towardsdatascience.com/interperable-vs-explainable-machine-learning-1fa525e12f48

Paul, A. M. (2010). *The Cult of Personality Testing: How Personality Tests Are Leading Us to Miseducate Our Children, Mismanage Our Companies, and Misunderstand Ourselves.* Retrieved from https://books.google.de/books?isbn=1451604068

Piano, S. L. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*(9), 1-7. https://doi.org/10.1057/s41599-020-0501-9

Polli, F. (2019, October 29). Using AI to Eliminate Bias from Hiring. *Harvard Business Review*. Retrieved from https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring

Ponce-López V., Chen, B., Oliu, M., Corneanu, C., Clapés, A., Guyon, I., . . . Escalera, S. (2016). ChaLearn LAP 2016: First Round Challenge on First Impressions – Dataset and Results. In G. Hua, & H. Jégou (Eds.), *Computer Vision – ECCV 2016 Workshops* (pp. 400-418). Cham, Switzerland: Springer.

Qin, R., Gao, W., Xu, H., & Hu, Z. (2016). Modern physiognomy: an investigation on predicting personality traits and intelligence from the human face. *Science China Information Science*, *61*(5), 1-27 https://doi.org/10.1007/s11432-016-9174-0

Raja, K., Ferrara, M., Franco, A., Spreeuwers, L., Batskos, I., Gomez-Barrero, F. D. W. M., ... & Busch, C. (2020). Morphing Attack Detection--Database, Evaluation Platform and Benchmarking. *IEEE Transactions on Information Forensics and Security. Manuscript submitted for publication.* Retrieved from https://arxiv.org/abs/2006.06458

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2012). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five Inventory-10 (BFI-10). *GESIS-Working Papers, 23,* 5-32. Retrieved from https://nbn-resolving.org/urn:nbn:de:0168-ssoar-312133

Rojas, M., Masip, D., Todorov, A., & Vitria, J. (2011). Automatic prediction of facial trait judgments: Appearance vs. structural models. *PloS one*, *6*(8), e23323. doi: 10.1371/journal.pone.0023323

Rothmann, S., & Coetzer, E. P. (2003). The big five personality dimensions and job performance. *SA Journal of Industrial Psychology*, *29*(1), 68-74. https://doi.org/10.4102/sajip.v29i1.88

Röber, T. E. (2021). Automated Personality Prediction Based On Facial Features. *Manuscript submitted for publication.*

Saucier, G., & Goldberg, L. R. (1996). Evidence for the Big Five in analyses of familiar English personality adjectives. *European journal of Personality*, *10*(1), 61-77. doi: 10.1002/(SICI)1099-0984(199603)10:1<61::AID-PER246>3.0.CO;2-D

Stidham, H., Summers, J., & Shuffler, M. (2018). Using the five factor model to study personality convergence on student engineering design teams. In D. Marjanović, M. Štorga, S. Škec, N. Bojčetić, & N. Pavković, *DS 92: Proceedings of the DESIGN 2018 15th International Design Conference* (pp. 2145-2154). Dubrovnik, Croatia: Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb. https://doi.org/10.21278/idc

Teijeiro-Mosquera, L., Biel, J. I., Alba-Castro, J. L., & Gatica-Perez, D. (2014). What your face vlogs about: Expressions of emotion and big-five traits impressions in YouTube. *IEEE Transactions on Affective Computing*, *6*(2), 193-205. doi: 10.1109/TAFFC.2014.2370044

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS ONE*, *14*(11), 1-20. https://doi.org/10.1371/journal.pone.0224365

Van den Berg, S. M. (2019). Generalized linear models: logistic regression. *Analyzing data using linear models* (pp.299-317). Enschede, Netherlands: University of Twente.

Van Noorden, R. (2020). The ethical questions that haunt facial-recognition research. *Nature, 587,* 354-358. https://doi.org/10.1038/d41586-020-03187-3

Ventura, C., Masip, D., & Lapedriza, A. (2017). Interpreting CNN Models for Apparent Personality Trait Regression. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1705-1713. doi: 10.1109/CVPRW.2017.217.

Ward, R., & Heys, T. (2020, July 10). Could Black Lives Matter speed up the introduction of ethnicity pay reporting? [Blog post]. Retrieved from https://www.peoplemanagement.co.uk/experts/legal/could-black-lives-matter-speed-up-introduction-ethnicity-pay-reporting

Wei, X. S., Zhang, C. L., Zhang, H., & Wu, J. (2017). Deep bimodal regression of apparent personality traits from short video sequences. *IEEE Transactions on Affective Computing*, *9*(3), 303-315. doi: 10.1109/TAFFC.2017.2762299

Zhang, T., Qin, R. Z., Dong, Q. L., Gao, W., Xu, H. R., & Hu, Z. Y. (2017). Physiognomy: Personality traits prediction by learning. *International Journal of Automation and Computing*, *14*(4), 386-395. https://doi.org/10.1007/s11633-017-1085-8

Zhang, D. D. (2013). *Automated biometrics: Technologies and systems.* Retrieved from https://books.google.de/books?isbn=1461545196