BRIEF REPORT

REVISED Prediction of self-efficacy in recognizing deepfakes based on personality traits [version 3; peer review: 2 approved]

Juneman Abraham[1], Heru Alamsyah Putra[1], Tommy Prayoga[2], Harco Leslie Hendric Spits Warnars[3], Rudi Hartono Manurung[4], Togiaratua Nainggolan[5]

[1]Psychology Department, Faculty of Humanities, Bina Nusantara University, Jakarta, 11480, Indonesia
[2]Content Collision, Jakarta, 11470, Indonesia
[3]Information System Concentration, Doctor of Computer Science Department, Bina Nusantara University, Jakarta, 11530, Indonesia
[4]Japanese Department, Faculty of Humanities, Bina Nusantara University, Jakarta, 11480, Indonesia
[5]Research Center for Social Welfare, Village, and Connectivity, National Research and Innovation Agency, Jakarta, 10340, Indonesia

v3 **First published:** 19 Dec 2022, **11**:1529
https://doi.org/10.12688/f1000research.128915.1
Second version: 10 Jul 2023, **11**:1529
https://doi.org/10.12688/f1000research.128915.2
Latest published: 12 Oct 2023, **11**:1529
https://doi.org/10.12688/f1000research.128915.3

Abstract

Background: While deepfake technology is still relatively new, concerns are increasing as they are getting harder to spot. The first question we need to ask is how good humans are at recognizing deepfakes - the realistic-looking videos or images that show people doing or saying things that they never actually did or said generated by an artificial intelligence-based technology. Research has shown that an individual's self-efficacy correlates with their ability to detect deepfakes. Previous studies suggest that one of the most fundamental predictors of self-efficacy are personality traits. In this study, we ask the question: how can people's personality traits influence their efficacy in recognizing deepfakes? **Methods:** Predictive correlational design with a multiple linear regression data analysis technique was used in this study. The participants of this study were 200 Indonesian young adults. **Results:** The results showed that only traits of Honesty-humility and Agreeableness were able to predict the efficacy, in the negative and positive directions, respectively. Meanwhile, traits of Emotionality, Extraversion, Conscientiousness, and Openness cannot predict it. **Conclusion:** Self-efficacy in spotting deepfakes can be predicted by certain personality traits.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 3 (revision) 12 Oct 2023	✓ view	
version 2 (revision) 10 Jul 2023	? view	✓ view
version 1 19 Dec 2022	? view	✗ view

1. **Sandra Grinschgl**, University of Graz, Graz, Austria

2. **Dilrukshi Gamage**, Tokyo Institute of Technology, Tokyo, Japan

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
deepfake detection, deepfake recognition, self-efficacy, personality, traits

This article is included in the Social Psychology gateway.

Corresponding author: Juneman Abraham (juneman@binus.ac.id)

Author roles: Abraham J: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Putra HA**: Data Curation, Formal Analysis, Investigation, Project Administration, Resources, Validation, Visualization, Writing – Original Draft Preparation; **Prayoga T**: Conceptualization, Formal Analysis, Methodology, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Warnars HLHS**: Formal Analysis, Methodology, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation; **Manurung RH**: Data Curation, Formal Analysis, Methodology, Resources, Validation, Visualization, Writing – Original Draft Preparation; **Nainggolan T**: Formal Analysis, Funding Acquisition, Methodology, Resources, Validation, Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This research was funded by DIPA of Directorate of Research, Technology, and Community Development, Directorate General of Higher Education, Research, and Technology, The Indonesian Ministry of Education, Culture, Research, and Technology, in accordance with the Research Contract for Fiscal Year 2022, Number: 454/LL3/AK.04/2022, dated 17 June 2022, assigned to Juneman Abraham as the Principal Investigator.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Abraham J, Putra HA, Prayoga T *et al.* **Prediction of self-efficacy in recognizing deepfakes based on personality traits [version 3; peer review: 2 approved]** F1000Research 2023, **11**:1529 https://doi.org/10.12688/f1000research.128915.3

First published: 19 Dec 2022, **11**:1529 https://doi.org/10.12688/f1000research.128915.1

REVISED **Amendments from Version 2**

The authors add the directions of the hypotheses, conclusion, and implications of this present study.

Any further responses from the reviewers can be found at the end of the article

Introduction

One of the biggest threats and disruptions to privacy and democracy in this digital age is deepfake technology. A 'deepfake' or synthetic media, is a video editing technology that manipulates and mimic a person's facial expressions, mannerisms, voice, and inflections based on a large amount of data of other people to create a hyper-realistic video depicting them doing or saying things that never happened (Westerlund, 2019).

The current consensus is that the average human's ability in recognizing deepfakes is similar to the machines (Vitak, 2022). However, the result seems to vary depending on their own confidence and belief in their cognitive abilities. Some studies suggest that some individual differences determine if a person is good at recognizing deepfakes or not (Shahid *et al.*, 2022). In this study, we will look at the relationship between personality traits and people's self-reported efficacy in recognizing deepfakes.

The HEXACO personality model describes six facets of personality structures: Honesty-humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to experience (Lee & Ashton, 2009; unpublished report; Zettler *et al.*, 2020). This instrument model is selected because of three reasons: (1) the measurement covers a wider and more complex range of personality facets that go beyond the five-factor model (Ashton & Lee, 2007); (2) the Honesty-Humility factor measures traits like sincerity, boastfulness, pretentiousness, and fair-mindedness that are associated with dishonest or inauthentic behaviors (Ashton & Lee, 2008) relating to self-reported efficacy; and (3) the model provides flexibility in measuring contextually unique situations (Oostrom *et al.*, 2019; Pletzer *et al.*, 2020). Advancements in information technology, including AI, socially intelligent robots, and other autonomous systems, will have a profound impact on human life, necessitating research in typical personality to understand and address individual differences in adapting to these new challenges (Matthews *et al.*, 2021), not to mention deepfakes. Multiple studies in various contexts have shown that personality traits influence an individual's self-efficacy (Lodewyk, 2018).

The Honesty-humility dimension reflects an individual's fair-mindedness, modesty, and cooperation. A person with high Honesty-humility might not think they are good at recognizing deepfakes, regardless of their true ability while an individual with low Honesty-humility might be biased in their self-reported ability in recognizing a deepfake.

Emotionality reflects an individual's degree of anxiousness, fearfulness, and sentimentality - the experience of anxiety in response to life's stressors. To overcome this anxiety, the sense of being able to recognize deepfakes is important to reduce that anxiety. One way to become less anxious is to appreciate deepfakes as a "cultural technology" (Cover, 2022) that contains artistic and creative values. People with high Emotionality may be more motivated to use deepfakes as an "antidote" from the pressures of everyday life, so they have higher self-reported efficacy to detect them, not to be avoided but as potential things to be used according to their interests (technology appropriation; see Prayoga & Abraham, 2017)

Extraversion reflects an individual's degree of sociability. Individuals high in Extraversion might have higher self-efficacy due to their higher social esteem, boldness, and familiarity. Van der Zee *et al.* (2002) found that extroverts are friendly and less formal in their interactions with others. This is closely connected with emotion recognition (part of emotional intelligence) which affects the success of negotiations. By using the paradigm of the social construction of technology (Kwok & Koh, 2021), humans are parties who "negotiate" with technology to better recognize the technology, including deepfakes, and can adapt it to not become victims of technology—or misappropriate technology for evil interests—but rather agents who utilize technology to improve humanity and prevent harm posed by technology (such as deepfakes).

An individual's degree of cooperation, tolerance, flexibility, and patience is reflected in the Agreeableness dimension. More agreeable people are at a larger risk for security, and social engineers (like deepfake designers) specifically target Agreeableness attributes like benevolence and compliance.

Conscientiousness reflects precisions, cautiousness, and a degree of self-control. Individuals with higher Conscientiousness thread might have higher self-efficacy in recognizing deepfakes. This is in line with the hypothesis of Köbis *et al.* (2021) that increasing Conscientiousness will make people motivated to invest cognitive resources to detect deepfakes, thereby enhancing their capacity to recognize truth and decreasing their desire to spread false information.

Openness reflects the willingness to experience new things and is associated with lower risk aversion. Research by Uebelacker and Quiel (2014) shows that open people don't create suitable coping mechanisms because they misjudge their vulnerability to being a target of social engineering (like deepfake designers).

This confirmatory study tested the hypotheses that the dimensions of HEXACO personality traits, i.e. (1) Honesty-humility (H), (2) Emotionality (E), (3) Extraversion (X), (4) Agreeableness (A), (5) Conscientiousness (C), and (6) Openness (O) can predict self-reported efficacy in recognizing deepfakes. H, A, and O traits would predict the efficacy in negative directions, while E, X, and C would predict it in positive directions.

The entire HEXACO model offers a thorough picture of a person's personality traits. The dynamics of the six qualities may provide a multidimensional perspective on one's cognitive, emotional, and behavioral responses to deceptive information when taking into account its ability to predict self-reported efficacy in spotting deep fakes. For instance, people with higher levels of Conscientiousness may investigate media with great care, whereas people with higher levels of Agreeableness may be less sceptic and more trusting. Based on the predictive powers of the six qualities, HEXACO may be able to shed light on a person's overall sense of vulnerability or resilience to digital deception.

Methods

Ethical considerations

This present study was initially approved by the Bina Nusantara University Research Committee, vide Letter of Approval No. 042/VR.RTT/VI/2021, strengthened with Letter No. 127/VR.RTT/VI/2022. The ethical decree is stated in Article 1 Paragraph 2 of the Letter.

Written informed consent was obtained from all participants of this study, which included consent for the research procedure to be carried out and for the publication of this article containing anonymized, analyzed, and interpreted data.

Participants filled out an electronic survey questionnaire consisting of demographic data and two scales, namely HEXACO Personality Traits (as the predictors) and Self-efficacy in recognizing deepfake (as the criterion variable). The design of this study was predictive correlation. There is only one data collection stage. There is no exposure in this study because the research was not an experimental study.

The eligibility criteria of the samples were young adults aged 18–25 years (Generation Z), which, according to a YouGov survey, is an age group who are concerned about a deepfake video of themselves going viral online (Help Net Security, 2022; unpublished report). In addition, Generation Z account for more than a quarter, precisely 26.47% of the total Indonesia's population (Badan Pusat Statistik, 2020a, 2020b). This group were the less likely to risk falling victim to misinformation like deepfakes compared to the older generation (Caramancion, 2021). The 18 to 24 age group was the most confident one in detecting deepfakes (iProov, 2020). Thus, understanding the self-efficacy of this age group in relation to their individual differences provides a huge potential for deepfakes detection strategies.

The participants of this study were 200 young adults (139 women, 61 men; M=22.06 years old; SD=1.98 year) who came from a non-Western country, Indonesia, and were recruited using a convenience sampling technique. The number of sample came from a calculation using the Sample Size Calculator (Calculator.net, 2022), with the following parameters: Confidence level of 95%, population size of 71,509,082 and population proportion of 26.47% - which was the total population of generation Z in Indonesia, as well as a margin of error of 6.2% - which is still in the range of 3–7%, the acceptable one (National Institutes of Health, 2005; unpublished report).

The research was conducted for 6 months from planning, participant recruitment, to data analysis. The research location is in Indonesia in an online setting for 3 months, namely 1 May to 31 July 2022. The research was a cross-sectional study, so no follow-up procedure was applied.

To measure self-efficacy in recognizing deepfakes, the authors constructed a self-efficacy measuring tool based on Bandura's theory (1977) which was adapted with the recommended checklists to pay attention when detecting deepfakes from the cyber-security company Norton taken from its unpublished report (Johansen, 2020). The introductory question was: "How sure are you that you can recognize or detect the presence of non-original or unnatural or unnatural elements (*e.g.* because it has been EDITED/MANIPULATED) from every image, photo, sound, and video you encounter?" Examples of items were: (1) I feel able to see abnormal eye movements; (2) I feel that I recognize awkward faces, *e.g.* if someone's face is pointing in one direction and the nose is pointing the other way; (3) I feel able to see any inappropriate skin tone in a video; (4) I am confident of being able to recognize when a person's face does not seem to convey the emotion that should be in line with what the person is supposed to say. There were six answer choices, ranging from "Feeling Very Incompetent" (scored 1) to "Feeling Very Capable" (scored 6).

To measure personality traits, this study used the short version of HEXACO-PI-R (60 items) (Lee & Ashton, 2009) with a scoring key. The response option ranged from "Strongly Disagree" (scored 1) to "Strongly Agree" (scored 6). The author translated the measuring tool into Indonesian.

All psychological scales in the questionnaire were tested for validity and reliability with the criteria of item validity (corrected item-total correlation) of at least 0.250 and internal consistency (Cronbach's α) of at least 0.600. A number of HEXACO trait items were eliminated because they did not meet these criteria. The test results are listed in Table 1.

The underlying data (Abraham & Alamsyah, 2022a), complete questionnaire (Abraham & Alamsyah, 2022b), and analysis script (Abraham, 2023) are openly available.

Multiple regression, when examining the hypothesized relationship between HEXACO traits and self-reported efficacy in identifying deep fakes, offers several advantages over simple correlational analysis. It enables the simultaneous examination of all traits, determining each trait's unique contribution while controlling for others. This method also filters out shared variance, detects multicollinearity, and allows for predictive model development. Additionally, multiple regression can discern the relative importance of each predictor, and overall, provides a more nuanced and precise assessment of how HEXACO traits jointly influence the self-efficacy to detect deep fakes.

Table 1. Descriptives (*N*=200).

Variable	Cronbach's α	Corrected Item-Total Correlations	*n* of items [before; after validation]	*M*	*SD*	*SE*
Honesty-humility	0.851	0.534-0.723	10; 6	2.910	1.015	0.072
Emotionality	0.671	0.433-0.528	10; 3	2.680	0.952	0.067
Extraversion	0.760	0.363-0.811	10; 5	2.325	0.766	0.054
Agreeableness	0.698	0.305-0.533	10; 6	3.782	0.653	0.046
Conscientiousness	0.817	0.478-0.625	10; 6	2.664	0.894	0.063
Openness	0.729	0.472-0.619	10; 5	2.702	0.819	0.058
Self-efficacy in recognizing deepfake	0.935	0.483-0.696	23; 23	4.360	0.762	0.054

Note. M = mean, *SD* = standard deviation, *SE* = standard error.

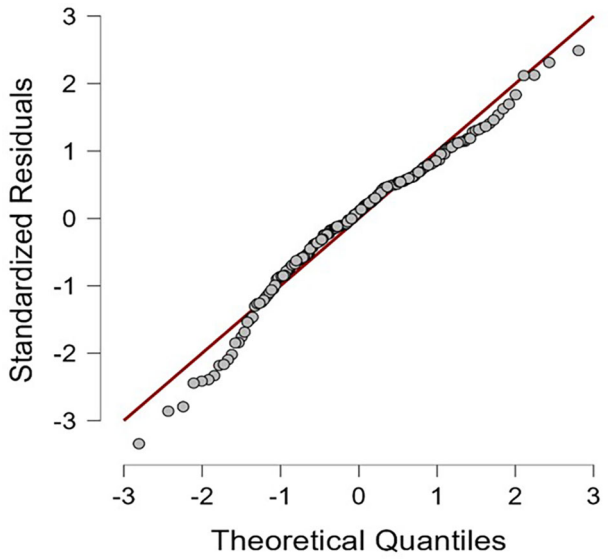

Figure 1. Normal probability (Q-Q) plot of multiple regression model's standardized residuals.

Results

Demographically, some participants were residents of DKI Jakarta province (*N*=90) which is the capital of Indonesia. In addition, other participants were residents of the Java Island (non-DKI Jakarta; *N*=86); Sumatera Island (*N*=21); and the rest (*N*=3) came from East Kalimantan, North Maluku, and West Nusa Tenggara provinces.

The psychometric properties and descriptive statistics of the variables are shown in Table 1. The results of this study indicate that the residuals of the multiple regression model are normally distributed (Figure 1) and all HEXACO personality dimensions are negatively correlated with self-efficacy in recognizing deepfakes; except for Agreeableness, which positively correlated (see Table 2). However, the results of the regression analysis with $F(6,199)=13,295$, $p=0.000$, $R^2=0.292$, showed that only Honesty-humility and Agreeableness were able to predict the efficacy (see Table 3). No difference was found between women and men, $t(198)=-0.120$, $p=0.904$, Cohen's $d=0.018$, *SE* Cohen's $d=0.154$, in terms of self-efficacy.

Table 2. Pearson's Correlations (*N*=200).

Variable		1	2	3	4	5	6	7
1. H	Pearson's *r*	—						
	p	—						
2. E	Pearson's *r*	0.641***	—					
	p	1.524e-24	—					
3. X	Pearson's *r*	0.510***	0.378***	—				
	p	1.178e-14	3.511e-8	—				
4. A	Pearson's *r*	-0.487***	-0.548***	-0.364***	—			
	p	2.469e-13	4.219e-17	1.132e-7	—			
5. C	Pearson's *r*	0.740***	0.606***	0.554***	-0.443***	—		
	p	6.084e-36	1.965e-21	1.668e-17	5.348e-11	—		
6. O	Pearson's *r*	0.674***	0.591***	0.460***	-0.483***	0.641***	—	
	p	7.910e-28	3.221e-20	7.048e-12	4.106e-13	1.713e-24	—	
7. SE	Pearson's *r*	-0.463***	-0.367***	-0.285***	0.465***	-0.403***	-0.381***	—
	p	5.244e-12	9.018e-8	4.268e-5	4.229e-12	3.278e-9	2.591e-8	—

Note. H = Honesty-humility, E = Emotionality, X = Extraversion, A = Agreeableness, C = Conscientiousness, O = Openness, SE = Self-efficacy in recognizing deepfake, *r* = Pearson's correlation coefficient, *p* = statistical significance of observed results.
*p <0.05,
**p <0.01,
***p <0.001.

Table 3. Multiple linear regression predicting self-efficacy in recognizing deepfake (*N*=200).

							Collinearity Statistics	
Model		***B***	***SE***	***β***	***t***	***p***	**Tolerance**	***VIF***
H_0	(Intercept)	4.360	0.054		80.871	3.532e-154		
H_1	(Intercept)	3.733	0.491		7.603	1.234e-12		
	Honesty-humility	-0.192	0.077	-0.255	-2.491	0.014	0.349	2.863
	Emotionality	0.023	0.070	0.029	0.332	0.740	0.475	2.104
	Extraversion	0.003	0.075	0.003	0.044	0.965	0.655	1.528
	Agreeableness	0.361	0.088	0.309	4.090	6.326e-5	0.643	1.555
	Conscientiousness	-0.068	0.085	-0.079	-0.798	0.426	0.372	2.691
	Openness	-0.026	0.083	-0.028	-0.312	0.755	0.462	2.166

Note. B = unstandardized beta, *SE* = standard error for the unstandardized beta, β = the standardized beta, *p* = statistical significance of observed results, *VIF* = variance inflation factor (*VIF* < 4 reflects no multicollinearity).

Table 3 shows the unadjusted (*B*) and adjusted (β) estimates for each predictor of which the potential confounders are the personality traits dimensions other than the focused predictor.

Discussion

Recognizing all deep fakes elements requires a certain level analytical capability and general intelligence (Ahmed, 2021). We need to look not just at people's cognitive abilities, but also at their belief in carrying out these abilities to recognize the information of deep fakes contextually. In other words, their self-efficacy.

This study found that, to a certain degree, individuals' personality traits do affect their self-efficacy in terms of detecting deepfakes. Because self-efficacy expression depends on context-to-context, it is not surprising that some traits can predict it better than the others.

Personality trait of Honesty-humility had negative predictive correlation with self-efficacy in recognizing deepfakes, β=-0.255, $t(193)$=-2.491, p<0.05 (Table 3). "Persons with very high scores on the Honesty-Humility scale avoid manipulating others for personal gain, feel little temptation to break rules, are uninterested in lavish wealth and luxuries, and feel no special entitlement to elevated social status" (Lee & Ashton, 2009, para 1). A person's Honesty-humility trait do not want to engineer others but, ironically, this trait makes them vulnerable to being manipulated by others (Ternovski *et al.*, 2021), including deepfakes. It can drive higher errors for the trait in recognizing deepfakes, exposing weaknesses that could be exploited.

Thompson *et al.* (2016, p. 54) once stated, "Honesty-Humility may not only be less likely to exploit others, they *may* also be strongly opposed to being the target of exploitation." However, this study found that when faced with deepfakes, Generation Z individuals high in Honesty-Humility feel overwhelmed and struggle to identify these manipulations.

That is a notable discovery of this present study, and could be explained by the findings of Weger *et al.* (2022) that Honesty-Humility has a negative correlation with general (r=-0.168, p=0.002) and specific (r=-.0270, p<0.001) technology acceptance. This is reinforced by the findings of Sindermann *et al.* (2020) that Honesty-Humility has a negative correlation with all aspects of technology acceptance, namely perceived usefulness (r=-0.25, p<0.001), perceived ease of use (r=-0.16, p<0.001), intention to use (r=-0.17, p<0.001), and predicted usage (r=-0.18, p<0.001). In fact, someone with high technology affinity is able to perceive deepfakes less negatively (Kleine, 2022). This is presumably because they feel they have knowledge and "master" deepfakes.

Therefore, to not fall for deepfakes, Generation Z with a high Honesty-Humility trait need to reduce their conservative attitude towards technology in order to detect potential harm and even utilize deepfakes effectively. Future studies can test this with an experimental design that involves measuring these two traits and people's ability to detect malicious vs. non-malicious deepfakes videos.

Not as hypothesized, Emotionality cannot predict self-efficacy in recognizing deepfakes, β=-0.029, $t(193)$=0.332, p>0.05 (Table 3). Austin and Vahle (2016) found that Emotionality—a trait that is positively correlated with empathy and social engagement—can predict the dimensions of Enhance (providing support and reassurance as interpersonal emotion management strategies) and Divert (the practice of using humor and pleasure pursuits to lift the spirits of others) of the Managing the Emotion of Others Scale (MEOS). This means that the Emotionality dimension is also positively correlated with the emotional intelligence needed to recognize deepfakes. Yang *et al.* (2022) emphasized the pivotal role of emotional intelligence in improving artificial intelligence technology so that it becomes a useful deepfake in the context of clinical encounters. By knowing that deepfakes themselves are increasingly being prepared with elements of emotional intelligence, then recognizing deepfakes also requires a better one; and this intelligence can actually be found in people with higher Emotionality. However, individuals high in Emotionality might be less confident in their own ability to accurately recognize deepfakes, as they might consider more factors and doubt themselves more (Thompson, 1998). With this uncertain direction, it is not surprising that no predictive power of Emotionality was found on self-efficacy.

Extraversion is a personality trait that cannot predict self-efficacy in recognizing deepfakes, β=0.003, $t(193)$=0.044, p>0.05 (Table 3). Hosler *et al.* (2021) put forward that detecting deepfakes is actually recognizing unnatural displays of emotion in voices and faces. Emotion apparently plays a central role in recognizing deepfakes because emotion is a higher-level semantic construct—which is difficult to counterfeit up to now—that could offer hints for detection. In an unpublished report, Kill states that emotion recognition is an ability that is honed in someone with a high extraversion trait (2021). However, Extraversion is also found to be positively correlated with excitement-seeking and a lower preference for consistency (Uebelacker & Quiel, 2014) - whereas "pairwise self-consistency learning" (Zhao *et al.*, 2021, p. 15023)

is needed to recognize deepfakes. Therefore, the positive and negative predictive powers of Extraversion on deepfake detection seem to neutralize each other.

Agreeableness trait can predict self-efficacy in recognizing deepfakes; however, not as hypothesized, the direction was found positive – not negative, β=0.309, t(193)=4.090, p<0.05 (Table 3). People with high Agreeableness are eager to cooperate and reach a compromise with others (Lee & Ashton, 2009). One of the good "others" in the context of deepfake recognition or detection is the "wisdom of the crowds" (Groh *et al.*, 2022), which Surowiecki (2004) defines as "the collective intelligence that arises when our imperfect judgments are aggregated". Agreeing with (or high Agreeableness to) the collective intelligence should reduce the chance of falsely recognizing deepfakes, including its algorithm attempts that present visual obstructions such as misalignment, partial occlusion, and inversion.

Agreeableness trait has a positive correlation with perception of forensic science (Sarki & Mat Saat, 2020). Deepfakes detection can be seen as part of forensic science. People with high agreeableness are known for their cooperativeness; agreeableness is often referred to as safeguards against antisocial behavior (Frias Armenta & Corral-Frías, 2021), including - in the context of this present study - deepfakes creation and distribution. They esteem innovative forensic methods in their environment and have a positive attitude toward it for the common good (Sarki & Mat Saat, 2020).

People who are more agreeable tend to make more accurate decisions about whether to believe information, which reduces their vulnerability to victimization (Cho *et al.*, 2016). This is confirmed by the empirical findings of van Winsen (2020) that agreeable individuals exhibit more secure online behavior and have a lesser likelihood of becoming a victim of cybercrime.

This study found that Conscientiousness was not able to predict self-efficacy in recognizing deepfakes, β=-0.079, t(193)=-0.798, p>0.05 (Table 3). Although deepfake recognition requires conscientious characteristics such as prudence and a sense of responsibility, Lawson and Kakkar's (as cited in Sütterlin *et al.*, 2022) research recently found that Conscientiousness is partially correlated with belief in conspiracy and conservatism - making it less efficacious in recognizing deepfakes.

This study found that Openness was not able to predict self-efficacy in recognizing deepfakes β=-0.028, t(193)=-0.312, p>0.05 (Table 3). In an unpublished report, Jin (2020) found that values of Openness to change do not correlate with the perceived ethical implications of deepfakes (*e.g.*, "These videos can uncontrollably deceive and influence many people", p. 24). In addition, contrast with the certain direction of the influence of Agreeableness and Honesty-humility on the self-efficacy; the direction of the Openness prediction is ambiguous. On the one hand, Openness is related to the low ability to recognize deepfakes. It is because Openness was found to be positively correlated with cognitive ability (Curtis *et al.*, 2015; Rammstedt *et al.*, 2016), but cognitive abilities encourage more protective online behavior, indicated by more interest in discussing how people who use deepfakes manipulate their audiences - rather than developing ability to apply scepticism on the authenticity of videos (Ahmed, 2021). On the other hand, there is a logic in favor of Openness as a buffer to prevent vulnerabilities from being manipulated by social engineering. For example, Eftimie *et al.* (2022) associated Openness with cognitive exploration tendencies which, based on their study, will stimulate responsible behavior including security best practices - which in the context of this study is deepfake recognition.

Based on the study findings, to avoid falling for deep fakes, there are two "optimal" personality traits that are worth exercising, i.e. Honesty-Humility and Agreeableness. *First*, the Honesty-Humility trait needs to be positioned strategically so that people with this trait can not easily be trapped or "absorbed" by the counterfeits from deepfakes technology, ie by reducing conventionalism (Leone *et al.*, 2012) towards technology, that is allegedly inherent in this trait. *Second*, agreeableness trait should be directed at various deepfake detection methods and technologies that are beneficial to community members.

A number of studies have shown that both general and technological self-efficacy are able to predict the actual ability associated with the use of the technology (Alnoor *et al.*, 2020; Raghuram *et al.*, 2003; Tetri & Juujärvi, 2022). This is because the efficacies determine organizing actions, behavioral intention and strategies, and preparedness for change, as well as reducing emotional sensitivity which is a source of performance anxiety.

Of course, there is no denying the possibility of inflated or overestimated belief, or the Dunning-Kruger effect (Koc *et al.*, 2022), which in the context of this study means that people who have high self-efficacy in detecting deepfakes actually have low actual abilities. In their research on bullshit detection, Cavojová *et al.* (2022) explained that the overestimation is caused by *metacognitive (un) awareness*, i.e. "These highly overconfident people suffer from a double curse – not only they do not know, but they also do not know that they do not know ... [that] is the result of self-enhancement motivation" (p. 1, 2).

To conclude, an individual's self-efficacy in spotting deepfakes is significantly predicted by personality traits, especially Honesty-Humility and Agreeableness. Higher Agreeableness encourages greater recognition, however higher Honesty-Humility may make one more vulnerable to such manipulations because of its negative link with technology acceptance. Furthermore, other traits, such as extraversion, conscientiousness, and openness, did not significantly predict self-efficacy in spotting deepfakes. Contradictory factors neutralized Extraversion's impacts, Conscientiousness' anticipated advantages were eclipsed by beliefs linked with the trait, and Openness, despite its association with cognitive capacity, had no direct effect on the effectiveness of deepfake recognition. These results imply that understanding these interactions between personality traits, technology acceptance, and self-efficacy is essential for developing successful defenses against deepfakes.

The limitation of this research is the use of non-probability sampling with limited generalizability. Nevertheless, this study has implication for the development of psychoinformatics - a branch of psychology that explains attitudes, competencies, and behavior in using information technology. Further research is suggested to implement random sampling and experimental methods to ensure a causal–not only predictive–relationship between personality traits and deepfakes detection self-efficacy.

Data availability

Underlying data

Zenodo: Dataset of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits. https://doi.org/10.5281/zenodo.7357400 (Abraham & Alamsyah, 2022a).

The project contains the following underlying data:

- Dataset of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits.xlsx (Raw data)

Extended data

Zenodo: Questionnaire of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits. https://doi.org/10.5281/zenodo.7413517 (Abraham & Alamsyah, 2022b).

The project contains the following extended data:

- Questionnare-HEXACO and Self-efficacy.docx

Zenodo: Analysis Script of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits. https://doi.org/10.5281/zenodo.8111881 (Abraham, 2023).

The project contains the following extended data:

- Analysis Script of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits.jasp (Analysis script)

References

Abraham J: **Analysis Script of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits. [Analysis script].** *Zenodo.* 2023;
Publisher Full Text

Abraham J, Alamsyah H: **Dataset of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits. [Data set].** *Zenodo.* 2022a.
Publisher Full Text

Abraham J, Alamsyah H: **Questionnaire of Prediction of Self-efficacy in Recognizing Deepfake based on Personality Traits. Zenodo. [Extended data].** 2022b.
Publisher Full Text

Ahmed S: **Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes.** *Personal. Individ. Differ.* 2021; **182**: 111074.
Publisher Full Text

Alnoor AM, Al-Abrrow H, Abdullah H, *et al.*: **The impact of self-efficacy on employees' ability to accept new technology in an Iraqi university.**

Glob. Bus. Organ. Excell. 2020; **39**(2): 41–50.
Publisher Full Text

Ashton MC, Lee K: **Empirical, theoretical, and practical advantages of the HEXACO model of personality structure.** *Pers. Soc. Psychol. Rev.* 2007; **11**(2): 150–166.
PubMed Abstract | Publisher Full Text

Ashton MC, Lee K: **The HEXACO model of personality structure and the importance of the H factor.** *Soc. Pers. Psychol. Compass.* 2008; **2**(5): 1952–1962.
Publisher Full Text

Austin EJ, Vahle N: **Associations of the Managing the Emotions of Others Scale (MEOS) with HEXACO personality and with trait emotional intelligence at the factor and facet level.** *Personal. Individ. Differ.* 2016; **94**: 348–353.
Publisher Full Text

Bandura A: **Self-efficacy: Toward a unifying theory of behavioral change.** *Psychol. Rev.* 1977; **84**(2): 191–215.
PubMed Abstract | Publisher Full Text

Caramancion KM: **The demographic profile most at risk of being disinformed.** *Proceedings of the 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS).* 2021, May; 1–7.
Publisher Full Text

Cavojová V, Šrol J, Brezina I: **Why people overestimate their bullshit detection abilities: Interplay of cognitive factors, self-esteem, and dark traits.** *PsyArXiv.* 2022: 1–35.
Publisher Full Text

Cho JH, Cam H, Oltramari A: **Effect of personality traits on trust and risk to phishing vulnerability: Modeling and analysis.** *Proceedings of the 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA).* 2016, Mar; 7–13.
Publisher Full Text

Cover R: **Deepfake culture: The emergence of audio-video deception as an object of social anxiety and regulation.** *Continuum.* 2022; **36**(4): 609–621.
Publisher Full Text

Curtis RG, Windsor TD, Soubelet A: **The relationship between Big-5 personality traits and cognitive ability in older adults – a review.** *Aging Neuropsychol. Cognit.* 2015; **22**(1): 42–71.
PubMed Abstract | Publisher Full Text

Eftimie S, Moinescu R, Răcuciu C: **Spear-phishing susceptibility stemming from personality traits.** *IEEE Access.* 2022; **10**: 73548–73561.
Publisher Full Text

Frias Armenta M, Corral-Frías NS: **Positive university environment and agreeableness as protective factors against antisocial behavior in Mexican university students.** *Front. Psychol.* 2021; **12**: 662146.
PubMed Abstract | Publisher Full Text | Free Full Text

Groh M, Epstein Z, Firestone C, *et al.*: **Deepfake detection by human crowds, machines, and machine-informed crowds.** *Proc. Natl. Acad. Sci.* 2022; **119**(1): e2110013119.
PubMed Abstract | Publisher Full Text | Free Full Text

Hosler B, Salvi D, Murray A, *et al.*: **Do deepfakes feel emotions? A semantic approach to recognizing deepfakes via emotional inconsistencies.** *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* 2021; 1013–1022.
Publisher Full Text

Koc E, Yurur S, Ozsahin M: **Problem-solving abilities of managers: Inflated self-efficacy beliefs.** *J. Hosp. Tour. Insights.* 2022.
Publisher Full Text

Köbis NC, Doležalová B, Soraperra I: **Fooled twice: People cannot detect deepfakes but think they can.** *IScience.* 2021; **24**(11): 103364.
PubMed Abstract | Publisher Full Text | Free Full Text

Kleine F: *Perception of deepfake technology - The influence of the recipients' affinity for technology on the perception of deepfakes [master's thesis].* Dieburg, Germany: Mediencampus of Hochschule Darmstadt; 2022.
Reference Source

Kwok AO, Koh SG: **Deepfake: A social construction of technology perspective.** *Curr. Issue Tour.* 2021; **24**(13): 1798–1802.
Publisher Full Text

Leone L, Desimoni M, Chirumbolo A: **HEXACO, social worldviews and socio-political attitudes: A mediation analysis.** *Pers. Individ. Differ.* 2012; **53**(8): 995–1001.
Publisher Full Text

Lodewyk KR: **Associations between trait personality, anxiety, self-efficacy and intentions to exercise by gender in high school physical education.** *Educ. Psychol.* 2018; **38**(4): 487–501.
Publisher Full Text

Matthews G, Hancock PA, Lin J, *et al.*: **Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems.** *Pers. Individ. Differ.* 2021; **169**: 109969.
Publisher Full Text

Oostrom JK, de Vries RE, De Wit M: **Development and validation of a HEXACO situational judgment test.** *Hum. Perform.* 2019; **32**(1): 1–29.
Publisher Full Text

Pletzer JL, Oostrom JK, Bentvelzen M, *et al.*: **Comparing domain-and facet-level relations of the HEXACO personality model with workplace deviance: A meta-analysis.** *Pers. Individ. Differ.* 2020 Jan **152**;152: 109539.
Publisher Full Text

Prayoga T, Abraham J: **Technopsychology of IoT optimization in business world**.Lee I, editor. *The Internet of things in the modern business environment.* IGI Global; 2017; (pp. 50–75).
Publisher Full Text

Raghuram S, Wiesenfeld B, Garud R: **Technology enabled work: The role of self-efficacy in determining telecommuter adjustment and structuring behavior.** *J. Vocat. Behav.* 2003; **63**(2): 180–198.
Publisher Full Text

Rammstedt B, Danner D, Martin S: **The association between personality and cognitive ability: Going beyond simple effects.** *J. Res. Pers.* 2016; **62**: 39–44.
Publisher Full Text

Sarki ZM, Mat Saat GA: **Adaptability traits and perception of forensic science among Investigating Police Officers (IPOs) in Nigeria.** *Salus J.* 2020; **8**(1): 75–92.

Shahid F, Kamath S, Sidotam A, *et al.*: **"It matches my worldview": Examining perceptions and attitudes around fake videos.** *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems,.* 2022; **255**: 1–15.
Publisher Full Text

Sindermann C, Riedl R, Montag C: **Investigating the relationship between personality and technology acceptance with a focus on the smartphone from a gender perspective: results of an exploratory survey study.** *Future Internet.* 2020; **12**(7): 110.
Publisher Full Text

Surowiecki J: *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations.* Doubleday & Co.; 2004.

Sütterlin S, Lugo RG, Ask TF, *et al.*: **The role of IT background for metacognitive accuracy, confidence and overestimation of deep fake recognition skills**.Schmorrow DD, Fidopiastis CM, editors. *Lecture Notes in Computer Science (subseries Lecture Notes in Artificial Intelligence, Augmented Cognition).* Springer; 2022; **13310**: 103–119.
Publisher Full Text

Ternovski J, Kalla J, Aronow PM: **Deepfake warnings for political videos increase disbelief but do not improve discernment: Evidence from two experiments.** *OSF [Preprint].* 2021.
Publisher Full Text

Tetri B, Juujärvi S: **Self-efficacy, internet self-efficacy, and proxy efficacy as predictors of the use of digital social and health care services among mental health service users in Finland: A cross-sectional study.** *Psychol. Res. Behav. Manag.* 2022; **15**: 291–303.
PubMed Abstract | Publisher Full Text | Free Full Text

Thompson RA: **Emotional competence and the development of self.** *Psychol. Inq.* 1998; **9**(4): 308–309.
Publisher Full Text

Thompson M, Carlson D, Hunter E, *et al.*: **We all seek revenge: The role of honesty-humility in reactions to incivility.** *J. Behav. Appl. Manag.* 2016; **17**(1): 50–65.

Uebelacker S, Quiel S: **The social engineering personality framework.** *Proceedings of the 2014 Workshop on Socio-Technical Aspects in Security and Trust.* 2014, July; 24–30.
Publisher Full Text

Van der Zee K, Thijs M, Schakel L: **The relationship of emotional intelligence with academic intelligence and the Big Five.** *Eur. J. Personal.* 2002; **16**(2): 103–125.
Publisher Full Text

Weger K, Easley T, Branham N, *et al.*: **Individual differences in the acceptance and adoption of AI-enabled autonomous systems.** *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* 2022, Sep; **66**(1): 241–245.
Publisher Full Text

Westerlund M: **The emergence of deepfake technology: A review.** *Technol. Innov. Manag. Rev.* 2019; **9**(11): 39–52.
Publisher Full Text

van Winsen B: *Determining secure digital behavior of individuals using HEXACO personality traits [master's thesis].* Rotterdam, Netherlands: Erasmus School of Economics; 2020.
Reference Source

Yang HC, Rahmanti AR, Huang CW, *et al.*: **How can research on artificial empathy be enhanced by applying deepfakes?** *J. Med. Internet Res.* 2022; **24**(3): e29506.
PubMed Abstract | Publisher Full Text | Free Full Text

Zhao T, Xu X, Xu M, *et al.*: **Learning self-consistency for deepfake detection.** *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV).* 2021; 15023–15033.
Publisher Full Text

Zettler I, Thielmann I, Hilbig BE, *et al.*: **The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation.** *Perspect. Psychol. Sci.* 2020; **15**(3): 723–760.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 3

Reviewer Report 13 December 2023

https://doi.org/10.5256/f1000research.157245.r214616

✔ **Sandra Grinschgl**
Institute of Psychology, University of Graz, Graz, Austria

The authors revised their manuscript based on my suggestions. I still think, however, that some of the interpretations and conclusions could be softened and put in less causal way.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Human-technology interaction, personality psychology, cognitive psychology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 14 September 2023

https://doi.org/10.5256/f1000research.152590.r185694

✔ **Dilrukshi Gamage**
Department of Innovation Science, Tokyo Institute of Technology, Tokyo, Japan

In summary, authors have used HEXACO Personality Traits as the independent variable and the criteria used in Self-efficacy in recognizing deepfake. The 6 factors of the HEXACO were used to understand the self efficacy of recognizing deepfakes in a 6 point likert scale. I believe the

dependent was the self-efficacy in recognizing deepfakes, where the authors constructed a self-efficacy measuring tool based on Bandura's theory. And this was more adopted using the Notion company criteria.

Overall, authors have significantly improved the article, in terms of the clarity in explaining the methods and evaluations.

I would recommend not to explain the methods first with "There is only one data collection stage. There is no exposure in this study because the research was not an experimental study." Although as a reviewer I know you may be addressing a comment, but if this is published, the readers have no clue on it. Explain the method first and then to be clear its a survey questionnaire .

I would also recommend to add a conclusion part to the manuscript summarizing the key takes aways - why some traits were removed and what implications does it reflect to the predicted efficacy. Also add the overall implication of such predictors to the efficacy items.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational social science, deepfakes social implications

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 17 July 2023

https://doi.org/10.5256/f1000research.152590.r185695

?

Sandra Grinschgl
Institute of Psychology, University of Graz, Graz, Austria

This is a revised version of a manuscript that I previously reviewed. The authors addressed my previously raised comments and improved their manuscript.

Here are a few comments that the authors might still want to consider:

1. The authors now included a brief paragraph on their hypotheses, however, to me those seem rather unspecific. In what direction would each personality trait predict self-reported efficacy in identifying deep fakes? How could the traits together predict variance in self-reported efficacy in identifying deep fakes? A related question: Why did the authors not have hypotheses for the correlational analyses?

2. Some of the interpretations in the discussion seem a bit too extreme and could be downscaled. For instance: "Generation Z with trait Honesty-Humility feels helpless, so it is less functional in detecting deepfakes." I don't think the authors can draw this conclusion as

they did not measure the actual ability to detect deepfakes.

"Therefore, the effects of Extraversion traits appear to cancel out of each other resulting in no predictive correlation with the self-efficacy" -> this should be formulated softer as it's only a suggestion.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Human-technology interaction, personality psychology, cognitive psychology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Version 1

Reviewer Report 20 February 2023

https://doi.org/10.5256/f1000research.141554.r162152

Dilrukshi Gamage
Department of Innovation Science, Tokyo Institute of Technology, Tokyo, Japan

Overall, the objective of this brief report is to find out the personality traits that affect the efficacy of spotting deepfakes.

- **Is the work clearly and accurately presented and does it cite the current literature?**

Since some of the literature cited are unpublished reports, I am not sure of the credibility. Since those were absorbed to the main study, this is pretty questionable.

- **Is the study design appropriate and does the work have academic merit?**

As mentioned in the introduction, the authors have taken the personality traits HEXACO from an unpublished report in 2009. I am curious what the authors consider other personality traits and also why did not cite any peer reviewed article with verified and validated factors. On the other hand, introduction does not provide a smooth understanding to why such model was selected and how others have conducted such explanations.

- **Are sufficient details of methods and analysis provided to allow replication by others?**

Methods were illustrated in a very awkward pattern, for example at once I was not sure doing such a sample size for what- the method of data collection, it took a while to understand that this is a questionnaire, and authors' explanation of other statements made this bit complicated than a straight forward mention.

The survey questionnaire was unclear - as I understand it's a 6 point Likert scale, but I am not

sure if the authors show any deepfake video before the question asked, and if so what are those.

- **If applicable, is the statistical analysis and its interpretation appropriate?**

The authors were not clear upfront about their analysis structure or procedure. I believe this was not pre-registered as well. In the results Authors explain the regression analysis and also correlation coefficient of the variables (personality traits), but due to the fact that this was not explain or mentioned as hypothesis, it is not clear the objective of the results and the strategy it provide to show the evidence.

- **Are all the source data underlying the results available to ensure full reproducibility?**

Data seems to be available.

- **Are the conclusions drawn adequately supported by the results?**

The authors do not provide any concrete conclusions as per se but, the discussion is somewhat leaning toward their conclusions.

But I have many issues with the way the study is conducted and how authors could claim statement as follow - “Honesty-humility trait do not want to engineer others but, ironically, this trait makes them vulnerable to being manipulated by others (Ternovski et al., 2021), including deepfakes, especially in the context of political greediness.” ---> how did you think of political context, because the study did not support any assumption? Did you ask this question?

Is the work clearly and accurately presented and does it cite the current literature?
No

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
No

Are the conclusions drawn adequately supported by the results?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational social science, deepfakes social implications

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 10 February 2023

https://doi.org/10.5256/f1000research.141554.r160499

?

Sandra Grinschgl
Institute of Psychology, University of Graz, Graz, Austria

Summary: This article deals with a highly relevant topic - the identification of deepfakes. The authors investigated potential predictors of the self-reported ability to detect deepfakes, namely individuals' personality traits based on the HEXACO model. While the traits honesty-humility and agreeableness were indeed shown to be predictors for self-reported ability to identify deepfakes, emotionality, extraversion, conscientiousness and openness were not. A higher honesty-humility was related to a lower self-reported ability to detect deepfakes whereas a higher agreeableness was related to a higher self-reported ability to detect deepfakes.

I think this is an interesting article on a very acute topic, however, I think it would benefit from some revisions. I outline my concerns below.

Major comments:

- Regarding the clear and accurate presentation of the authors' work: In my opinion, the introduction is lacking an elaborate justification for testing the HEXACO traits as potential predictors of the (self-reported) ability to detect deepfakes. Why might especially those traits act as predictors? The authors could, for instance, argue that typical personality traits might also be relevant when it comes to the application of other artificial intelligence technologies (e.g., robots). For a reference on this behalf see Matthews et al. (2021)[1].
- The introduction does not end with clear hypotheses, thus, to me it is unclear whether this research was exploratory or confirmatory.
- Table 1: Why was the number of items reduced for the HEXACO traits? What does "after validation" refer do? This choice of method might potentially influence the interpretation of results.
- I think the discussion would benefit from a paragraph that focuses on the bigger picture in which the results can be placed. For instance, it could be discussed what conclusions we can draw from the study findings when it comes to the increasing distribution of deepfakes. What might be the "optimal" personality to not fall for deepfakes? Also, I think the authors should discuss how self-reported abilities in detecting deepfakes might be related to actual abilities in detecting deepfakes. This seems especially relevant as individuals' self-estimation is not always accurate (e.g., above-average effect, Dunning-Kruger effect).

Minor comments:

- Page 3: "The current consensus is that the average human's ability in recognizing deepfakes is similar to the machines (Vitak, 2022)." I'm not sure what this exactly means. Does it mean that artificial intelligence technologies analyzing videos/pictures are as good as humans in identifying deepfakes from real videos?
- Only within the methods section it became fully clear to me that the authors did not actually measure participants' ability to detect deepfakes but rather participants' self-reported ability

to do so. I would recommend the authors to refer to “self-reported efficacy in recognizing deepfakes” already from the beginning of their article (and in the abstract) to avoid any confusion.
- Figure 1: I guess this graph shows the residuals for the multiple regression model. Please clearly state this in the figure description and corresponding manuscript text.
- Please report an effect size with the t-Test statistics on gender differences.
- I think especially the discussion could need some language-editing. Every paragraph starts with “This study found” which does not induce an optimal reading flow.
- Regarding reproducibility: Data of this study are already openly available (great!). I would suggest that the authors also make their analyses scripts available.

References

1. Matthews G, Hancock P, Lin J, Panganiban A, et al.: Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Personality and Individual Differences*. 2021; **169**. Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Human-technology interaction, personality psychology, cognitive psychology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.