Predicting Personality Traits from Hand-Tracking and Typing Behavior in Extended Reality under the Presence and Absence of Haptic Feedback

Jonathan Liebers
HCI Group
University of Duisburg-Essen
Essen, NRW, Germany
jonathan.liebers@uni-due.de

Felix Bernardi
HCI Group
University of Duisburg-Essen
Essen, NRW, Germany
felix.bernardi@stud.uni-due.de

Alia Saad
HCI Group
University of Duisburg-Essen
Essen, NRW, Germany
alia.saad@uni-due.de

Lukas Mecke
LMU Munich
Munich, BY, Germany
lukas.mecke@ifi.lmu.de

Uwe Gruenefeld
HCI Group
University of Duisburg-Essen
Essen, NRW, Germany
uwe.gruenefeld@uni-due.de

Florian Alt
LMU Munich
Munich, BY, Germany
florian.alt@ifi.lmu.de

Stefan Schneegass
HCI Group
University of Duisburg-Essen
Essen, NRW, Germany
stefan.schneegass@uni-due.de

Abstract

With the proliferation of extended realities, it becomes increasingly important to create applications that adapt themselves to the user, which enhances the user experience. One source that allows for adaptation is users' behavior, which is implicitly captured on XR devices, such as their hand and finger movements during natural interactions. This data can be used to predict a user's personality traits, which allows the application to accustom itself to the user's needs. In this study (N=20), we explore personality prediction from hand-tracking and keystroke data during a typing activity in Augmented Virtuality and Virtual Reality. We manipulate the haptic elements, i.e., whether users type on a physical or virtual keyboard, and capture data from participants on two different days. We find a best-performing model with an R^2 of 0.4456, with the error source stemming from the manifestation of XR , and that the hand-tracking data contributes most of the prediction power.

CCS Concepts

• **Human-centered computing** → *Empirical studies in HCI*; *Mixed / augmented reality*; *Virtual reality*.

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720270

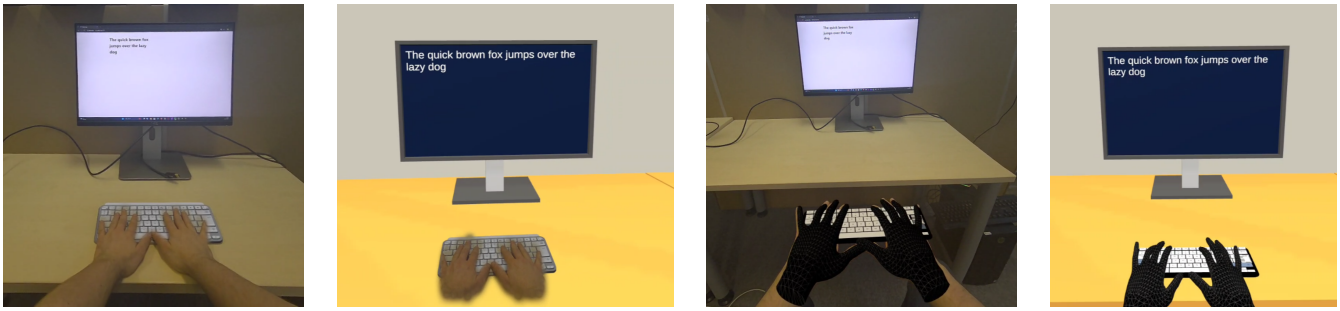
Keywords

Personality Prediction, HEXACO, Augmented Virtuality, Virtual Reality, Extended Reality, Hand Tracking

ACM Reference Format:
Jonathan Liebers, Felix Bernardi, Alia Saad, Lukas Mecke, Uwe Gruenefeld, Florian Alt, and Stefan Schneegass. 2025. Predicting Personality Traits from Hand-Tracking and Typing Behavior in Extended Reality under the Presence and Absence of Haptic Feedback. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3720270

1 Introduction & State of the Art

With Extended Reality (XR) increasingly becoming a ubiquitous part of daily life, it becomes essential for applications (apps) and developers to understand and have knowledge of their users. By knowing their users, apps can allow for adaptability, meaning that they can customize themselves to the user and react to their needs. Such adaptability allows for personalization and affects user experience [1, 20]. One particularly valuable aspect to know about users is their personality and associated personality traits. Personality and a person's behavior are two closely linked properties [23]. By leveraging a person's behavior, an application can predict their user's personality traits [23]. Vice versa, it has also been observed that personality traits predict smartphone usage, as in frequency and duration of actual behavior [24]. Personality can virtually influence many facets of interaction in Human-Computer Interaction (HCI) [17], ranging from the responses to advertisements [2, 8, 14], the time spent gaming [6], social dynamics [7], or to the interpretation and usage of emojis [25]. Therefore, knowing the user's personality is crucial to enable personalization and increase user experience, while it remains an underexplored area of HCI [22].



(a) AV-H. (b) VR-H. (c) AV-NH. (d) VR-NH.

Figure 1: Screencaptures from the Head-Mounted Display (HMD) illustrating the four conditions of the study. In (a) and (b), participants type on a physical keyboard with haptic feedback (H) positioned in front of them. They could see it directly through colored passthrough (a) or blended into the virtual environment (b). In (c) and (d), participants interacted with a virtual keyboard and received no haptic (NH) feedback. It was blended into the passthrough stream (c) or positioned in the virtual environment (d).

Knowledge about the user's personality can mainly be obtained from two sources. First, the user can be explicitly asked to complete a questionnaire, and their subjective feedback contains the information. While such a questionnaire can be embedded in an app, this is a precise but tedious way of obtaining this information, as it bothers the users and requires their cooperation. However, personality traits can also be predicted from the user's behavioral patterns that can be implicitly captured in an app [21]. For example, while a user naturally interacts with an app (e.g., on a touchscreen), their behavioral data is captured and can be used to predict their personality [11, 23]. Such knowledge of users' personalities is practically available for free, as it is based on actions that the user carries out anyway and can be elicited as a byproduct, as it is common in other domains of HCI [9].

One such behavior that can be used to elicit data for personality prediction is typing [3]. Typing is an activity that is frequently used and widely supported across many computing devices since users often need to input text. To the best of our knowledge, personality prediction was not yet explored through (a) typing in XR, and (b) by facilitating hand-tracking data. Therefore, this is important for numerous reasons. First, typing in XR can be different from typing in reality, as users can interact with (a) a virtual keyboard that provides no haptic feedback or (b) with a physical keyboard that they can feel and that provides haptic feedback. Second, an increasing number of XR devices allows for natural interaction through hand-tracking, meaning that users can directly interact with keyboards without the need to use dedicated hand-held controllers. This hand-tracking data, which can be captured at zero cost in addition to the actual keystrokes from the virtual or physical keyboard, provides an important opportunity to predict personality.

Therefore, we explore the possibilities of personality prediction in XR through a user study (N=20) with two sessions per participant conducted on different days. We use the HEXACO personality model [13]. There, we explore users' typing behavior when confronted with a virtual keyboard that does not provide haptic feedback and a physical keyboard that does provide haptic feedback. In addition, we ask the participants the NASA raw Task-Load Index (rTLX) questionnaire for their subjective feedback on perceived workload to correlate their responses with the personality prediction errors. We conduct this exploration in Augmented Virtuality (AV)[1] and VR as two manifestations of XR. We posit two research questions:

> **RQ1.** Can we predict users' personality traits from their typing interactions in XR (i.e., in AV and VR)?
> **RQ2.** How is this prediction influenced by the presence and absence of Haptics and the manifestation of XR?

To the best of our knowledge, we are the first to explore personality prediction across XR during typing, using hand-tracking and keystroke dynamics in conjunction with the presence and absence of haptics. We explore five different feature sets and find positive R^2 scores across all conditions for the best features, with the highest result yielding an R^2 of 0.4456, which is considered a "large effect" [5], and an average Root Mean Square Error (RMSE) of 0.36. Furthermore, we explore the sources of prediction errors following participants' subjective feedback and which features contribute most to the prediction of personality. There, we see that the hand-tracking data is the primary contributor to the predictive capability for personality traits, far surpassing the contribution of the keyboard interaction data (i. e., the actual typing behavior). Overall, personality prediction is feasible using widely available Head-Mounted Displays (HMDs), and the hand-tracking data contributes an essential amount of information, while the manifestation of XR is important.

2 Experiment

We conducted a lab-based user study to understand how well personality can be predicted from typing (keystroke) and hand-tracking data. In addition, we examine what factors influence the personality prediction error.

[1]We chose Augmented Virtuality as a close manifestation to Augmented Reality (AR), as our user study requires a single device that can also provide an immersive Virtual Reality (VR) experience with hand- and keyboard-tracking. A single device is important to avoid introducing confounding variables. Currently, to the best of our knowledge, these capabilities can only be provided by the Meta Quest 3 that utilizes a colored passthrough mode, falling into the class of Augmented Virtuality on Milgram's continuum [15].

2.1 Study Design

We explored two independent variables in our within-subject experiment. The first independent variable is *XR*, which refers to the visualization of the Virtual Environment (VE) in the HMD. It has two levels: *Augmented Virtuality (AV)* and *Virtual Reality (VR)*. AV refers to participants being able to view the real environment as it was displayed to them through the passthrough mode of the HMD, and VR refers to them being in an immersive VE. The second independent variable is *Haptics*, which, again, has two levels: *presence-of-haptics* ("H") and *absence-of-haptics* ("no haptics" – "NH"). Figure 1 provides an overview of our 2×2 design. To capture participants' personality traits, we use the HEXACO-60 questionnaire and calculate the six corresponding scores: "Honesty-Humility" (H), "Emotionality" (E), "Extraversion" (X), "Agreeableness" (A), "Conscientiousness" (C), and "Openness to Experience" (O). As dependent variables, we collected participants' hand-tracking and keystroke data with which we performed a regression that yields the participants' RMSE, i.e., the error between the regression result and participants' previously given answers to HEXACO-60. For this, our user study encompasses two sessions. Participants were invited to our lab twice on different days, and we used their first session data for model training and the second session data for model testing. This assures that the used data stems from different days, increasing the study's validity, as the model cannot rely on odd or outlier behavior from one session (e.g., the HMD being worn in a tilted or otherwise odd way).

2.2 Apparatus

Our apparatus consists of a Meta Quest 3 HMD, an application implemented in Unity3D, in addition to a python-based keystroke logger, a monitor, and a Bluetooth-connected physical keyboard. The python-based application is the central element of the apparatus. It controls the study state by providing information to the HMD what VE it should load, and it controls the execution of the repetitions. While it runs on a desktop computer, it also collects the keystrokes from the physical keyboard, for which we chose a Logitech MX Keys Mini, as it connects to it via Bluetooth. Furthermore, the python application keeps the typing state in a buffer and displays the typed keystrokes on a regular PC monitor to the participants, which is important for conditions AV-H and AV-NH, as participants would see the monitor through the Quests' colored passthrough mode. It also provides this information to the VR monitor.

Next, the capabilities of the Meta Quest 3 are used to capture participants' hand and finger movements through its hand-tracking capabilities. When Augmented Virtuality (AV) (i.e., AV-H or AV-NH) would be used, the Quest would turn on its passthrough functionality so that participants could see their surroundings. When VR (i.e., VR-H or VR-NH) would be used, they would be placed in an immersive replica of the lab. In the former case, participants would see the real monitor, while in the latter case, they would see a virtual representation of the monitor (cf., Figure 1). Participants would also type on a virtual keyboard in these conditions without haptics, namely AV-NH and VR-NH. The virtual keyboard is a size-wise exact replica of the Logitech MX Keys Mini. The state of this keyboard is sent to the python application through HTTP requests so that the python logger logs the virtual keystrokes similar to the logging of the attached Bluetooth keyboard in AV-H and VR-H. The hand-tracking data was logged directly on the Quest into a sequence of Tabulator-separated Values (TSV) files. Finally, we match the keystroke data obtained from the python application and the hand-tracking data obtained from the Quest through timestamp-based matching.

The physical keyboard, a Logitech MX Keys Mini, was used for the study because it is supported by Meta for keyboard-tracking. In the VR-NH condition, participants are in VR, but they type on a keyboard that is tracked and rendered into their VE. For this, we utilized Meta's keyboard tracking SDK[2], and the Logitech MX Keys Mini is one of the few supported models. The keyboard itself has a QWERTY-layout and a size of $296 \times 21 \times 132$mm. Notably, we keep the dimensions of the keyboard and its angle identical in all conditions. Thus, the virtual keyboard is the exact same size as the physical keyboard. Furthermore, the virtual keyboard is angled as if it would lie flat on a table. We keep these parameters constant across all conditions in order to avoid any induced confounding variables.

2.3 Power Analysis

To determine the number of participants to be invited, we performed a power analysis using G*Power (version 3.1.9.7). We requested the a priori required sample size given α, power, and effect size for a repeated-measures within-factors ANOVA. Given effect size $f = .3$, $\alpha = .05$, $\beta = .80$, number of groups = 1, and number of measurements = 4, G*Power suggested a total sample size of 17 ($\lambda = 12.24, F = 2.7981$).

2.4 Participants

We recruited 20 volunteers (two female, 17 male, and one preferred not to disclose) for the lab study via our institution's mailing lists and social media. We exceeded the number of suggested participants to fully balance the Latin Square. Participants were between 20 and 36 years old (M = 25.90, SD = 5.27). 16 were right-handed, and four were left-handed. They were asked to self-describe their VR usage through the agreement with the statement "I have lot of experience with VR" (L1) and "I use VR very often before taking part in the study" (L2) on a 7 pt. Likert scale (1 = "strongly disagree", 7 = "strongly agree"). The median response to L1 was 3 (IQR: 2), and to L2, they responded at a median of 3 (IQR: 3).

2.5 Procedure & Ethics

At first, participants were welcomed to our lab, and the experimenter explained the procedure to them. All questions that participants had were fully answered. Next, their written and informed consent was collected. Participants were assured that they could cancel their participation in the study at any point in time without any detriments. The research project and procedure received prior clearance from our institution's ethics committee. Then, we collected their HEXACO-60 scores using the respective questionnaires [13].

[2]Meta Quest Help. "Set up a tracked keyboard for Meta Quest", https://www.meta.com/help/quest/articles/headsets-and-accessories/meta-quest-accessories/tracked-keyboards-meta-quest, last retrieved on March 11, 2025.

Table 1: Test results as R^2 scores and mean RMSE values (denoted as $\overline{RMSE}$) of the Random Forest regressor and all tested features and all conditions. A better regression model yields a higher R^2 score, whereas a lower RMSE is favored. The RMSE is a per-participant metric and shows the deviation of the predictions towards a participant's reference scores. The values in brackets denote RMSE standard deviation. Concatenating all features yields the highest mean R^2 score across conditions and indicates the best model (Nr. 5). For R^2, the ideal score is 1.0, representing a perfect fit, while negative values suggest a model performing worse than simply predicting the mean.

	AV-H		AV-NH		VR-H		VR-NH	
Features	R^2	$\overline{RMSE}$	R^2	$\overline{RMSE}$	R^2	$\overline{RMSE}$	R^2	$\overline{RMSE}$
1. Flight Time (FT)	-.0016	.49 (.23)	-.0858	.54 (.20)	-.0197	.52 (.19)	-.1673	.54 (.24)
2. Dwell Time (DT)	.0388	.51 (.16)	-.1537	.55 (.22)	.0274	.52 (.16)	-.0268	.53 (.19)
3. Flight Times + Dwell Times	.1665	.46 (.19)	.0016	.51 (.20)	.1642	.48 (.16)	.0333	.50 (.22)
4. Hand Tracking	.0081	.60 (.41)	.2082	.46 (.19)	.4431	.36 (.18)	.2807	.42 (.18)
5. FT + DT + Hand Tracking	.0239	.50 (.24)	.2157	.45 (.20)	.4456	.36 (.18)	.2941	.41 (.19)

Participants were introduced to the HMD and helped in adjusting the straps and inter-pupillary distance. Next, we tested the four conditions (AV-NH, AV-H, VR-NH, and VR-H, cf., Figure 1) in counterbalanced order for one training repetition and then six more logged repetitions. The order of conditions was randomized using a Latin square. The initial training repetition was employed to let participants accustom themselves to the virtual environment and to try out the respective keyboard. Then, the actual typing began, and participants had to enter the sentence "The quick brown fox jumps over the lazy dog", as it is one sentence that contains every character. In case of typing mistakes, participants had to correct their input using the backspace key and were given visual feedback in case they entered a wrong character. After each condition, we asked participants the rTLX [18, 19]. The full session took approximately one hour. Finally, participants were invited to participate in a second study on another day, following the same procedure.

The study took place in a lab environment where participants were seated in XR. We placed them in front of a table, where the physical keyboard was placed for the haptics conditions (i.e., in AV-H and VR-H, cf., Figure 1: (a) and (b)). For the conditions without the influence of haptics, participants moved back a little on the provided chair so that they would type in-air and their hands would not touch the desk (i.e., in AV-NH and VR-NH, cf., Figure 1: (c) and (d)). This way, the absence of haptics would be realized.

3 Analysis

We first create a regressor that predicts participants' HEXACO-60 scores from their hand-tracking and keystroke data. Next, participants' RMSE values are calculated that denote the regressor's prediction error as the resulting dependent variable. Finally, we investigate how the prediction errors are influenced by the independent variables *XR* and *Haptics*.

3.1 Data Set

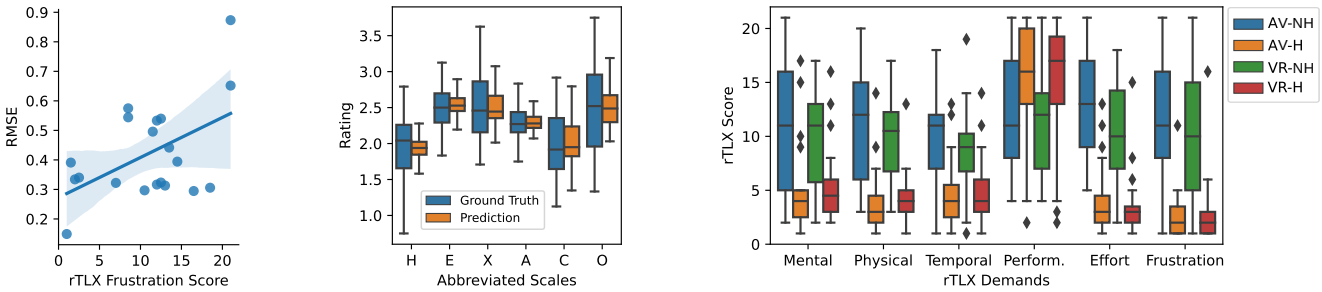
The data set consists of the hand-tracking data stream directly logged on the HMD and the keystroke events logged on the desktop computer. The former provides information about the spatiotemporal coordinates of the tracked hands and fingers of the wearer. These coordinates consist of each finger's positions ("pos.x", "pos.y", and "pos.z") and rotations ("rot.x", "rot.y", "rot.z", and "rot.w"). The hands are represented by the thumbs, index, middle, ring, and pinky finger, which are segmented into the fingertip, proximal, middle, distal, and metacarpal joints. Additionally, we logged the forearm stubs, the base of the hand, and the coordinates of the head, as provided by the HMD. In addition to the coordinates, a timestamp is present. Furthermore, the data set also consists of the actual keystrokes that participants entered on the keyboard, which is a series of individual key-up/key-down events, in addition to the typed string and timestamps. Therefore, the data set consists of, in total, 24 GB of data that is comprised of 960 samples = 20 Participants × 4 Conditions × 6 Repetitions × 2 Sessions. It is balanced in any regard. All samples obtained from the hand-tracking data stream are logged at a rate of 72 Hz. Since hand tracking can sometimes lose track under rare circumstances and produce "not a number" values (NaN), we apply linear interpolation to fill the gaps and match the keystroke and hand tracking data by timestamp.

3.2 Data Split

We strictly organize the data set given the participation in two sessions throughout a hold-out validation. Any model is always trained with the first session of the user study. The second session exists to test the model and to assess its performance under realistic circumstances. Therefore, we mimic a process where a model to predict participants' personalities is deployed on a device based on the first session's data and tested during usage on a second day.

3.3 Feature Engineering & Model Training

Feature Engineering. We create various features in our preprocessing from participants' keystrokes and hand-tracking data and train and test regression models with it. Every keyboard interaction (i. e., each trial) in the study resulted in a slightly different number of key presses, depending on whether a typing mistake was made. This also applies to the hand-tracking data, as every interaction resulted in a different number of captured frames, as there were slight variations in the time they took, even if no typing mistakes were made. This means that a varying degree of data is present in the time-axis for the analysis, yet its overall shape needs to be unified. The data is organized in a two-dimensional tabular format with shape $(n \times m)$, where the columns (n) denote the different features and the rows (m) correspond to the time axis. To unify the



(a) Correlation. (b) Ground Truth and Predictions. (c) rTLX responses.

Figure 2: In (a), the significant correlation between RMSE and frustration is shown. In (b), the ground truth is shown which is participants' responses to the HEXACO-60 questionnaire, corresponding to the training labels, together with the predictions of our best-performing regressor in condition "VR-H" for the features "FT + DT + Hand Tracking". Participants' rTLX responses are depicted in (c).

shape, we summarize each feature, i.e., each column in the data, using four aggregate functions (min, max, mean, and standard deviation). This means that we calculate these four aggregates for each column and concatenate them into one single-dimension feature vector for each trial that has a resulting size of $4 \times n$. This approach follows previous work that applied the same preprocessing method [12, 16].

First, from the keystroke data only, we calculate the flight and dwell times between key inputs and concatenate these into one respective single-dimension vector. The feature describing only the flight times between key presses is the first feature we explore and is listed as Nr. 1 in Table 1. Then, we also calculate the dwell times, i.e., how long each key press took, and again summarize them using the same four aggregate functions. We list it as Nr. 2 in Table 1. Next, we also concatenate the feature vectors of the flight times and dwell times. The results are listed as Nr. 3 in Table 1.

Then, we summarize the participants' hand-tracking data, which we obtained from the HMD, using the same aggregate functions. To avoid any participant-induced bias, we subtract all values in the sample from each feature's initial data point so that everything becomes relative to the beginning of the interaction. Per hand feature, i.e., per every single hand and finger's object's position and rotation, such as all included phalanges and bones, we again calculate the min, max, mean, and standard deviation and concatenate them into a one-dimensional vector. It particularly helps to unify the shape of the samples, as every interaction took a slightly different time. We list this data as Nr. 4 in Table 1. Finally, we concatenate all existing data, i.e., the dwell times, flight times, and hand tracking data, and list the results as Nr. 5 in Table 1.

Model Training. Next, we train a Random Forest regressor for each condition with the session 1 data and test it with the session 2 data. The training and testing labels are comprised of the HEXACO-60 scores that are calculated from the questionnaire, which is a vector of six float values, one corresponding to each dimension of the questionnaire [13]. The distribution of labels is visualized in Figure 2(b). Notably, the predictions of the regressor fall into a narrower range compared to the ground truth, indicating difficulties in predicting marginal values. For training the regressor, we exclusively use the scores and data that are obtained from the first session. Next, we test the regressor with the data obtained from the second session and compare it to the training scores. We calculate the R^2 metric to indicate each model's performance and list them in Table 1. Additionally, we calculate each participant's RMSE, i.e., the prediction error of the regressor, for every typing repetition they contributed to the study. Table 1 lists the participants' means and standard deviations for every condition. All in all, we avoid overfitting by strictly separating training and testing data, as both were elicited in the user study on different days. In particular, this means that participants had to take off the HMD between sessions so that the regressor cannot rely on the specifics of the interactions, such as an odd wearing style.

3.4 Factorial Analysis

Finally, we examine participants' RMSE in relation to the independent variables *XR* and *Haptics*. The RMSE is the prediction error, with lower values indicating more accurate predictions. To avoid inflating inferential statistics, we calculate the mean RMSE for each participant and condition, as multiple samples exist per participant and condition in the test dataset due to the six repetitions of the typing. Subsequently, we analyze these averaged values using a two-way repeated-measures analysis of variance (RM-ANOVA), where RMSE is the dependent variable. *XR* (with levels "AV" and "VR") and *Haptics* (with levels "H" and "NH") are the independent variables, following our 2×2 design.

4 Results

We find that the concatenation of all features, i.e., participants' flight times, dwell times, and hand tracking data, yields the highest mean R^2, as shown in Nr. 5 in Table 1. Overall, we find a mean R^2 score for "FT + DT + Hand Tracking" across all conditions of .2986, which is a large effect [5, p. 79 ff.]. The highest observed R^2 value corresponds to VR-H at 0.4456. Next, we explore what influences the root mean square errors for this best-performing model. The best-performing model's predictions using session 2 data, together

with the initial response of our participants (i.e., the training labels – ground truth), are visualized in Figure 2(b).

4.1 Two-Way RM-ANOVA on Root Mean Square Error

Assumptions. First, we check the assumptions of RM-ANOVA. Levene's test for the homogeneity of variances across groups (homoscedasticity) did not indicate that our groups have significant differences in variance ($F(3, 76) = .7512$, $p = .5248$). The assumption of sphericity is necessarily met due to the experiment design. We check if the RMSE is normally distributed and find that a Shapiro-Wilk test indicates that the data is not normally distributed ($W = .9247$, $p = .0002$). Therefore, we apply the aligned-rank transform [26].

Main Effects. First, we find a significant main effect for *XR*. VR (Med. = .31, IQR = .25) leads to significantly lower prediction errors compared to AV (Med. = .43, IQR = .31) with $F(1, 57) = 6.8080$, $p = .0116$, $\eta_p^2 = .1067$. This is confirmed by a post-hoc test ($t(57) = 2.6092$, $p = .0116$). However, a significant main effect could not be observed for *Haptics*, as "H" (Med. = .38, IQR = .31) did not show a significant difference compared to "NH" (Med. = .36, IQR = .25), $F(1, 57) = .1581$, $p = .6924$, $\eta_p^2 = .0028$. Also, the interaction effect between *XR* and *Haptics* did not show a significant difference ($F(1, 57) = 1.6403$, $p = .2055$, $\eta_p^2 = .0280$).

Contrasts. The pair-wise comparisons between conditions follow a p-value correction using Tukey's HSD. We find a significant difference between AV-H and VR-H with $t(57) = 2.8157$, $p = .0328$, as VR-H (Med. = .27, IQR = .23) lead to significantly lower prediction errors compared to AV-H (Med. = .44, IQR = .35). However, the contrast test for AV-H vs. AV-NH (Med. = .42, IQR = .25) did not show significant differences ($t(57) = .4022$, $p = .9778$). The contrast for AV-H vs. VR-NH (Med. = .35, IQR = .22) also was not observed to be significant, $t(57) = 1.2872$, $p = .5748$. Next, the contrast for AV-NH vs. VR-H showed a low p-value but was also not observed to be significant ($t(57) = 2.4135$, $p = .0859$). Furthermore, the comparison between AV-NH and VR-NH did not yield a significant difference ($t(57) = .8849$, $p = .8127$). Finally, the contrast of VR-H and VR-NH did not indicate a significant difference ($t(57) = -1.5285$, $p = .4275$).

4.2 Task-Load Index & Correlation to Prediction Error

We asked the participants for the rTLX after completing each condition in session 1. Their responses are visualized in Figure 2(c). It is evident that the workload primarily depends on the presence of haptics, as supported by another RM-ANOVA (see Appendix A). In summary, the presence of haptics leads to a significantly lower perceived workload compared to the absence of haptics. No significant difference could be observed for the influence of *XR* on participants' workload. We also observed that the reported frustration had a particularly high range of responses (cf., Figure 2(c)). A few participants mentioned during the study that they encountered difficulties interacting with the virtual keyboard, which they found to be small; however, other participants also rated low frustration values. To explore this further, we calculated a Pearson correlation between their self-reported frustration levels and their respective RMSE values. The results revealed a strong positive correlation (Pearson's $r(18) = .50$, $p = .0247$), as shown in Figure 2(a). This suggests that higher frustration may be associated with greater personality prediction error.

5 Discussion

Overall, we find that hand-tracking data contains most of the information necessary for personality prediction. This is particularly evident from Table 1, where the features associated with hand tracking exceed greatly. However, the keystroke features still contribute a small degree of information, as the concatenation of all data leads to the highest R^2. Yet, for the keystroke data alone, it is evident that they are not sufficient for reliable personality prediction, as their predictive power is very limited. Therefore, the utilization of hand-tracking data is necessary. Another limitation is that the regressor's predictions fall into a narrow range compared to the ground truth (cf., Figure 2(b)), indicating that personality traits at the extremes of the spectrum are not well predicted. It would be important to find a more fine-grained prediction approach so that all traits are well reflected across their full range, which is also demonstrated by the observed RMSE. Finally, it needs to be noted that our preprocessing treated all samples equally and did not account for typing mistakes.

Following our RM-ANOVA, we particularly find that the prediction error of personality depends on whether the behavior was sampled in AV or VR, which is primarily indicated by the significant main effect. Particularly, the prediction model for VR-H yields an R^2 of .4456, which denotes a large effect and that personality prediction is partially possible (RQ1), with the limitations regarding the range of predictions and their connected RMSE, as previously stated [5]. Also, it needs to be noted that the HEXACO-60 questionnaire also has a range into which participants' responses typically fall [13]. However, the AV-H model performs significantly worse, and the cause is connected to *XR* (RQ2), which requires further exploration. While it is apparent from the rTLX results (cf., Figure 2(c)) that AV-NH was rated at a higher workload and lower performance, the difference to VR-NH is small. We believe that the primary reason might be that participants saw the real environment through the colored passthrough, which could have influenced their behavior. This indicates that personality prediction might be easier in a more sterile, purely virtual environment that allows for less distraction and divergence in behavior compared to an AV environment. Additionally, we find that high frustration values significantly correlate with higher prediction errors. The reason for this might come from the decision to keep the dimensions of the virtual and physical keyboard the same; we did not change the size or its angle, as this would have introduced potential confounding variables. Therefore, our virtual keyboard is smaller than most of the ones used in existing applications, which might be considered a limitation, and one future research direction would be the inclusion of an oversized virtual keyboard. Also, we kept the visualization of the hands in AV-NH and VR-NH constant to avoid any further confounds. Thus, it is important to investigate further the incorporated factors for personality prediction, such as the workload associated with typing and hand-tracking and potentially associated dimensions related to keyboards in XR [4, 10].

6 Conclusion

In this study, we explored the influence of Extended Reality (XR) and particularly its manifestations of AV and VR on the ability to predict user's personality traits from their hand-tracking data during a typing activity. As a second factor, we investigated the presence and absence of haptics, i.e., whether the typing took place on a physical or virtual keyboard. Our results indicate that personality prediction is feasible in XR settings, with hand-tracking data yielding the most informative feature. The concatenation of hand-tracking and keystroke features produced the best results, particularly in virtual environments with haptics (VR-H). Our analysis revealed that prediction errors were significantly lower in VR settings compared to AV, suggesting that the immersive nature of VR facilitates the conveying of personality traits through possibly more stable behavior. Therefore, this work contributes to the field by performing one of the first explorations of such personality traits across multiple facets of XR, using hands and typing as the primary source of information. Future research could focus on improving prediction models, exploring additional features, and examining the impact of other environmental factors on user behavior in XR.

Acknowledgments

The presented work was funded by the German Research Foundation (DFG) under project number 521584224. We thank Marvin Prochazka for his help in creating the video.

References

[1] Ofer Arazy, Oded Nov, and Nanda Kumar. 2015. Personalityzation: UI personalization, theoretical grounding in HCI and design research. *AIS Transactions on Human-Computer Interaction* 7, 2 (2015), 43–69.

[2] Stefanos Balaskas, Aliki Panagiotarou, and Maria Rigou. 2023. Impact of Personality Traits on Small Charitable Donations: The Role of Altruism and Attitude towards an Advertisement. *Societies* 13, 6 (2023), 144. doi:10.3390/soc13060144

[3] Abeer A. N. Buker and Alessandro Vinciarelli. 2021. I Feel it in Your Fingers: Inference of Self-Assessed Personality Traits from Keystroke Dynamics in Dyadic Interactive Chats. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8. doi:10.1109/ACII52823.2021.9597389

[4] Francesco Chiossi, Yassmine El Khaoudi, Changkun Ou, Ludwig Sidenmark, Abdelrahman Zaky, Tiare Feuchtner, and Sven Mayer. 2024. Evaluating Typing Performance in Different Mixed Reality Manifestations using Physiological Features. *Proc. ACM Hum.-Comput. Interact.* 8, ISS, Article 542 (Oct. 2024), 30 pages. doi:10.1145/3698142

[5] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Routledge. doi:10.4324/9780203771587

[6] Lea C. de Hesselle, Dmitri Rozgonjuk, Cornelia Sindermann, Halley M. Pontes, and Christian Montag. 2021. The associations between Big Five personality traits, gaming motives, and self-reported time spent gaming. *Personality and Individual Differences* 171 (2021), 110483. doi:10.1016/j.paid.2020.110483

[7] Adam Goodkind. 2023. *Predicting Social Dynamics in Interactions Using Keystroke Patterns*. Ph. D. Dissertation. USA. Northwestern University.

[8] Jacob B. Hirsh, Sonia K. Kang, and Galen V. Bodenhausen. 2012. Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. *Psychological science* 23, 6 (2012), 578–581. doi:10.1177/0956797611436349

[9] Markus Jakobsson, Elaine Shi, Philippe Golle, and Richard Chow. 2009. Implicit Authentication for Mobile Devices. In *Proceedings of the 4th USENIX Conference on Hot Topics in Security (HotSec'09)*. USENIX Association, USA, 9.

[10] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3173574.3173919

[11] Ludwig Küster, Carola Trahms, and Jan-Niklas Voigt-Antons. 2018. Predicting personality traits from touchscreen based interactions. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–6. doi:10.1109/QoMEX.2018.8463375

[12] Jonathan Liebers, Sascha Brockel, Uwe Gruenefeld, and Stefan Schneegass. 2022. Identifying Users by Their Hand Tracking Data in Augmented and Virtual Reality. *International Journal of Human–Computer Interaction* (2022). doi:10.1080/10447318.2022.2120845

[13] Michael C. Ashton and Kibeom Lee. 2009. The HEXACO–60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment* 91, 4 (2009), 340–345. doi:10.1080/00223890902935878

[14] John C. Mowen, Eric G. Harris, and Sterling A. Bone. 2004. Personality traits and fear response to print advertisements: Theory and an empirical study. *Psychology & Marketing* 21, 11 (2004), 927–943. doi:10.1002/mar.20040

[15] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: a class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*, Hari Das (Ed.), Vol. 2351. SPIE, 282–292. doi:10.1117/12.197321

[16] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300340

[17] Kym E. Pocius. 1991. Personality factors in human-computer interaction: A review of the literature. *Computers in Human Behavior* 7, 3 (1991), 103–135. doi:10.1016/0747-5632(91)90002-I

[18] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. doi:10.1177/154193120605000909

[19] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*. Elsevier, Amsterdam, The Netherlands, 139–183. doi:10.1016/s0166-4115(08)62386-9

[20] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Quay Au, Bernd Bischl, Markus Bühner, and Heinrich Hussmann. 2019. 2. Opportunities and challenges of utilizing personality traits for personalization in HCI. In *Personalized human-computer interaction*, Mirjam Augstein, Eelco Herder, and Wolfgang Wörndl (Eds.). De Gruyter, Berlin and Boston, 31–64. doi:10.1515/9783110552485-002

[21] Albrecht Schmidt. 2000. Implicit human computer interaction through context. *Personal Technologies* 4, 2-3 (2000), 191–199. doi:10.1007/BF01324126

[22] Simarpreet Singh and Williamjeet Singh. 2024. AI-based personality prediction for human well-being from text data: a systematic review. *Multimedia Tools and Applications* 83, 15 (2024), 46325–46368. doi:10.1007/s11042-023-17282-w

[23] Clemens Stachl, Quay Au, Ramona Schoedel, Samuel D. Gosling, Gabriella M. Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, Heinrich Hussmann, Bernd Bischl, and Markus Bühner. 2020. Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences of the United States of America* 117, 30 (2020), 17680–17687. doi:10.1073/pnas.1920484117

[24] Clemens Stachl, Sven Hilbert, Jiew-Quay Au, Daniel Buschek, Alexander de Luca, Bernd Bischl, Heinrich Hussmann, and Markus Bühner. 2017. Personality Traits Predict Smartphone Usage. *European Journal of Personality* 31, 6 (2017), 701–722. doi:10.1002/per.2113

[25] Sarah Theres Völkel, Daniel Buschek, Jelena Pranjic, and Heinrich Hussmann. 2019. Understanding Emoji Interpretation through User Personality and Message Context. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, 1–12. doi:10.1145/3338286.3340114

[26] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

A Two-Way RM-ANOVA on rTLX

We conduct a two-way repeated-measures ANOVA on the rTLX scores (cf., Figure 2(c)) to understand the influence of the independent variables *XR* with levels "AV" and "VR", and *Haptics* with levels "H" and "NH" onto participants' perceived workload. For this, we create the per-participant mean workload by summarizing their responses to the different scales of the rTLX. We invert the performance scale so that lower values consistently indicate a lower workload and better performance. We find that a significantly lower workload is created when a physical keyboard is used ("H")

compared to a virtual keyboard ("NH"). A significant difference for *XR*, i.e., "AV" and "VR" could not be observed.

Assumptions. Levene's test for the homogeneity of variance across groups (homoscedasticity) did not indicate that our groups have significant differences in variance, F(3, 76) = 1.0263, p = .3858. Sphericity is again necessarily met by the experiment design. A Shapiro-Wilk test indicates that the rTLX data is not normally distributed (W = .9469, p = .0023). Therefore, we again apply the aligned-rank transformation [26].

Main Effects. We find a significant main effect for *Haptics*: "NH" (Med. = 12.33, IQR = 5.42) leads to higher rTLX scores compared to "H" (Med. = 5.83, IQR = 3.08) with F(1, 57) = 103.7581, p < .0001, η_p^2 = .6454. This is confirmed by a post-hoc test (t(57) = -10.1862, p < .0001). The *XR* main effect, however, was not observed to be significant (F(1, 57) = .6621, p = .4192, η_p^2 = .0115), as no significant difference was shown between the rTLX scores in VR (Med. = 9.25, IQR = 6.29) compared to AV (Med. = 8.67, IQR = 8.21). Also, The interaction effect between XR and Haptics did not show a significant difference (F(1, 57) = 1.7268, p = .1941, η_p^2 = .0294).

Contrasts. We find that all four related contrast tests support the significant main effect for *Haptics*, showing that "H" yields consistently lower rTLX workload scores compared to "NH". First, AV-H (Med. = 5.67, IQR = 2.88) leads to lower rTLX workload scores compared to AV-NH (Med. = 12.67, IQR = 5.33), t(57) = -8.0098, p < .0001. Next, AV-H also leads to significantly lower scores compared to VR-NH (Med. = 11.83, IQR = 3.88), t(57) = -6.8721, p < .0001. Furthermore, VR-H (Med. = 6.17, IQR = 2.88) yields significantly lower scores than AV-NH (t(57) = 7.7225, p < .0001). Finally, VR-H also shows significantly lower rTLX scores compared to VR-NH (t(57) = -6.5848, p < .0001). The difference between AV-H and VR-H was not observed to be significant (t(57) = -.2873, p = .9916). Similarly, the comparison between AV-NH and VR-NH was also not shown to be significant (t(57) = 1.1377, p = .6680). All p-values were adjusted using Tukey's HSD.