Characterizing Personality from Eye-Tracking: The Role of Gaze and Its Absence in Interactive Search Environments

Jiaman He*
RMIT University
Naarm/Melbourne, Australia
jiaman.he@student.rmit.edu.au

Marta Micheli*
University of Turin
Torino, Italy
marta.micheli@unito.it

Damiano Spina
RMIT University
Naarm/Melbourne, Australia
damiano.spina@rmit.edu.au

Dana McKay
RMIT University
Naarm/Melbourne, Australia
dana.mckay@rmit.edu.au

Johanne R. Trippas
RMIT University
Naarm/Melbourne, Australia
j.trippas@rmit.edu.au

Noriko Kando
National Institute of Informatics
Tokyo, Japan
kando@nii.ac.jp

Abstract

Personality traits influence how individuals engage, behave, and make decisions during the information-seeking process. However, few studies have linked personality to observable search behaviors. This study aims to characterize personality traits through a multimodal time-series model that integrates eye-tracking data and *gaze missingness*–periods when the user's gaze is not captured. This approach is based on the idea that people often look away when they think, signaling disengagement or reflection. We conducted a user study with 25 participants, who used an interactive application on an iPad, allowing them to engage with digital artifacts from a museum. We rely on raw gaze data from an eye tracker, minimizing preprocessing so that behavioral patterns can be preserved without substantial data cleaning. From this perspective, we trained models to predict personality traits using gaze signals. Our results from a five-fold cross-validation study demonstrate strong predictive performance across all five dimensions: Neuroticism (Macro F1 = 77.69%), Conscientiousness (74.52%), Openness (77.52%), Agreeableness (73.09%), and Extraversion (76.69%). The ablation study examines whether the absence of gaze information affects the model performance, demonstrating that incorporating missingness improves multimodal time-series modeling. The full model, which integrates both time-series signals and missingness information, achieves 10-15% higher accuracy and macro F1 scores across all Big Five traits compared to the model without time-series signals and missingness. These findings provide evidence that personality can be inferred from search-related gaze behavior and demonstrate the value of incorporating missing gaze data into time-series multimodal modeling.

CCS Concepts

• **Information systems → Users and interactive retrieval**.

Keywords

Eye Tracking, Personality Prediction, Interactive Search

1 Introduction

Consider the different ways people behave when searching for information. Some are easily distracted, while others display a strong sense of curiosity. Certain individuals prefer to skim ahead to the end, whereas others move back and forth through the material to piece together a deeper understanding. The motivations for searching also vary: some seek confirmation of what they already believe [8], while others are driven by the desire to learn something new [37]. Exploring these different behaviors is essential for advancing areas such as recommendation systems [15, 30, 55, 74], personalization [25, 34], understanding confirmation bias [53], or LLM-based simulations [21, 44, 47, 76]. Among the many factors that shape these behaviors, personality traits stand out as a foundational influence, impacting human patterns of thought, preference, and decision-making across contexts [38].

Personality traits fundamentally shape how and why people seek information. Curious and open individuals often seek intellectual engagement, while those high in neuroticism or intolerance of uncertainty tend to seek reassurance in negative contexts [32]. The Big Five model [52], which includes Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, is widely used in personalization research. Studies link personality to information seeking [24] and web search behaviors [4], showing its potential for personalized information retrieval (IR). However, while the importance of psychological dimensions in interaction and search has long been acknowledged as a major challenge in IR [5], the impact of the systematic exploration personality on IR behaviors and its data-driven modeling remains in its early stages.

A direction for modeling personality in IR lies in eye-tracking data. With mobile and wearable devices (e.g., Tobii Pro Glasses

3 [58], Apple Vision Pro [31]) now featuring eye-tracking capabilities [41], it is possible to collect such data in real-world settings. Eye movements provide insights into cognitive and attentional processes even before a user makes a decision [22]. Previous work has investigated links between personality and gaze patterns [6, 29], though studies focusing specifically on personality inference from eye movements during digital search interactions remain limited [12, 54, 72]. In addition, eye-tracking data also present challenges; for instance, *data missingness* occurs frequently when users are not looking at the screen, and noise often requires complex preprocessing. We address these issues by interpreting missingness–the absence of gaze–as informative signals and by using raw gaze data obtained directly from the eye tracker (Section 3.1.3), avoiding additional cleaning steps in order to capture behavioral patterns without extensive preprocessing.

Building on this perspective, we investigate the role of viewing behaviors in recognizing users' Big Five personality traits during search tasks. Our user study uses a GUI-based application displaying museum items, enabling the collection of rich data on interactive search behaviors, including gaze patterns and engagement dynamics. Our main contributions are as follows:

(1) We propose a multimodal time-series model that incorporates missing eye-tracking data as behavioral cues, integrating gaze and pupil signals to capture both attention patterns and underlying cognitive states.

(2) We show that users' Big Five personality traits can be predicted from eye-tracking data in complex search environments, achieving a classification performance across all five dimensions: Neuroticism (Macro F1 = 77.69%), Conscientiousness (74.52%), Openness (77.52%), Agreeableness (73.09%) (see Table 1).

(3) We present findings that characterize how users search and explore digitized museum collections, based on a controlled user study ($N = 25$) involving a graphical digital interface.

2 Related Work

2.1 Personality Traits and Search Behavior

Personality traits describe stable patterns in how individuals think, feel, and behave [36]. These traits shape emotional and motivational processes, influencing how people seek, process, and use information [62]. Prior research indicates that Big Five personality traits are associated with differences in information search behaviors [32].

Curiosity reflects a drive to explore novelty, complexity, or ambiguity [39]. It can be expressed as deprivation sensitivity, where individuals are motivated to close knowledge gaps, or as joyous exploration, where the act of learning itself is intrinsically rewarding [39]. By contrast, intolerance of uncertainty describes the tendency to experience uncertain situations as threatening, distressing, or undesirable [10].

Curiosity—particularly joyous exploration—is usually understood as part of openness/intellect, which captures differences in creativity, imagination, and intellectual engagement [39, 66]. Intolerance of uncertainty aligns more closely with neuroticism, reflecting vulnerability to negative affect and worry [5, 33]. Together with extraversion, agreeableness, and conscientiousness, these dimensions form the Big Five model of personality [49].

These traits play an important role in information-seeking behavior. Individuals high in joyous exploration may search broadly and openly, driven by intrinsic interest, while those high in deprivation sensitivity may pursue information more urgently to resolve perceived knowledge gaps. Conversely, individuals with high intolerance of uncertainty may engage in searching defensively, focusing on information that reduces ambiguity or confirms certainty, while avoiding open-ended exploration. In this way, personality traits can be seen as probabilistic predictors of how people approach information environments: shaping whether they explore widely, focus narrowly, avoid uncertainty, or embrace it.

From a dynamic perspective, personality traits should not be viewed as fixed responses but as distributions of tendencies across time and context [19]. For example, someone high in intolerance of uncertainty will not resist ambiguity in every instance but will do so more frequently and with greater intensity than someone lower on that trait. Similarly, individuals high in curiosity will tend to experience more frequent and stronger motivation to seek new information [32]. Thus, understanding personality traits provides valuable insight into the variability of information-seeking strategies across individuals.

In our study, we incorporate the Big Five personality traits [52], commonly summarized by the acronym **OCEAN**: Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N). We use eye-tracking data to examine how these traits relate to user behavior.

2.2 Eye Tracking and Personality Traits

Eye-tracking has been widely used to investigate people's cognitive states and their relationship to information-seeking behavior [13, 18, 22]. Research suggests that individuals with different personality traits may follow distinct eye movement patterns when searching for information [2]. More broadly, prior work has established that personality traits can influence gaze control and visual attention [2, 51, 59, 63]. This has motivated a series of studies exploring whether personality can be inferred directly from eye-tracking data.

Hoppe et al. [29] examined the Big Five traits during everyday activities. Participants wore head-mounted eye trackers while walking and shopping, and the resulting gaze data were used to train random forest classifiers. Their models achieved above-chance prediction of Extraversion, Neuroticism, Agreeableness, and Conscientiousness. Berkovsky et al. [6] extended this work to controlled settings, proposing a framework for predicting psychological characteristics—including the Big Five—based on passive viewing of images and videos. Using data from 21 participants, they applied supervised machine learning techniques and achieved encouraging levels of accuracy.

Other studies focused on interactive systems. Chen et al. [12] collected gaze data during product selection tasks with recommender systems, applying multiple classifiers and feature selection methods to predict personality. Millecamp et al. [54] studied gaze patterns in a music recommender system, with models predicting traits such as Openness to Experience, though their accuracy was not yet sufficient for practical applications.

Expanding to immersive contexts, Khatri et al. [40] examined user behavior in a virtual reality shop, incorporating three-dimensional gaze features to predict the Big Five. Woods et al. [72] investigated

social media use, tracking participants as they browsed their Facebook News Feeds. Using only 20 seconds of gaze data per user, classification produced acceptable results for Extraversion and Conscientiousness, though performance was weaker for other traits.

Collectively, these studies highlight the dual role of eye tracking: it not only provides insights into cognitive and information-seeking processes but also offers a potential signal for inferring personality. Personality-related differences in gaze behavior suggest that personality-aware models may help explain why individuals vary in how they explore, attend to, and process information.

A practical challenge is that eye-tracking data often contain noise, such as blinks or missing samples when users look away [35]. Previous studies have typically handled these gaps by interpolation, replacement, or removal [7, 20, 22, 50, 71]. However, we argue that instances of looking away are themselves informative, as they may reveal valuable aspects of human behavior. To leverage this, we propose a novel method that preserves these missing segments. Our approach, detailed in Section 3.2, integrates the missing data into a time-series framework to better capture behavioral patterns.

2.3 Time-Series Modeling and Missing Data

Cole et al. [14] analyzed time-series activity patterns to study search behavior. Similarly, eye-tracking data can be represented as a multivariate time series, capturing variables such as gaze coordinates, pupil size, and gaze velocity continuously over time [42]. At each timestep, a vector of measurements is recorded, and the sequence of these vectors traces the dynamics of visual attention as it evolves moment by moment.

Most prior work on personality inference from gaze has reduced these streams to aggregate features, such as fixation counts, blink rates, or average saccade lengths[6, 29, 54, 72]. While useful, such handcrafted summaries discard the fine-grained temporal patterns, such as prolonged fixations, systematic scanning, or rhythmic shifts in attention, that may carry diagnostic information about personality. By contrast, sequence models such as recurrent neural networks [26], convolutional temporal encoders [70], and more recently Transformers [45] are designed to capture dependencies across time, making them well suited for modeling how gaze evolves during complex tasks.

To motivate this shift, it is helpful to look at parallel domains where similar challenges arise. In human activity recognition (HAR), wearable sensors generate continuous multivariate streams (e.g., accelerometer, gyroscope) that are best understood as sequences [56]. Likewise, in healthcare, physiological signals such as ECG or PPG are modeled as multivariate time series to capture patterns in heart rhythms or respiration [28]. In both cases, moving from handcrafted features to sequence modeling has led to significant improvements in predictive accuracy and robustness [27]. These domains demonstrate the value of treating behavioral signals as structured temporal data—an approach we adopt for eye-tracking in information retrieval.

Another challenge with eye-tracking is the prevalence of missing data. Blinks, momentary loss of calibration, or glances away from the screen introduce gaps in the signal. Earlier studies often discarded these segments or filled them in with simple interpolation [29, 54]. However, research in healthcare time series has shown that missingness can itself be informative [46]: the duration and frequency of gaps may reflect underlying behavioral or physiological states. Modern approaches augment sequential inputs with masking vectors and temporal gap features, allowing models to distinguish between transient noise or short-term dropouts and more prolonged periods of missingness in the signal [11].

Despite the success of these methods in other domains, they have not been applied to eye-tracking for personality prediction. Our study builds on these insights by proposing a missing-data-aware sequence modeling framework, which treats gaze not only as a multivariate temporal signal but also as a behavioral record where the absence of data may carry meaningful information about individual differences in attention and personality.

3 Methodology

3.1 Experimental Setup

Experiments were conducted using the *Minpaku Guide* [65], an iPad application developed for the National Museum of Ethnology in Osaka, Japan. For this study, the English version of the app was employed. The application provides extensive content related to the museum's artifacts and is based on an *ostensive search model [9]*. The interface follows the same structural and functional principles as a search engine results page (SERP): users begin by browsing a grid of artifact photos, which they can select to view detailed information and scroll to explore related items [65]. The app consists of multiple page types, including (i) modular image grid (shown in Figure 1a), (ii) individual page, which contains description about a single museum object (shown in Figure 1b), and (iii) map-based views.

3.1.1 Procedure. The study followed a structured procedure: participants gave informed consent, completed a pre-task questionnaire and short training, performed the search task, and concluded with a post-questionnaire and brief interview with stimulus recall.

This paper uses the data collected in the following two phases:

(1) **Personality Assessment.** As part of pre-task questionnaire, participants completed the *BFI-44 questionnaire* [17] in English to measure their Big Five personality traits. This self-reporting instrument has been shown to be reliable [3].

(2) **Search Task.** Participants explored the Minpaku Guide and selected five items they liked most, marking them as *favorites*. No time limit was imposed to avoid biasing exploratory behavior. In addition to user's eye movement (see Section 3.1.3), the application also logged detailed interaction data, including item offsets and vertical scroll positions, since all the pages were scrollable.

This study was reviewed and approved by the Institutional Review Board (IRB) of National Institute of Informatics (NII), Japan.

In preparing the application for the task, we ensured that the quality and arrangement of the images on the first page were appropriate for the experiment. This page, which participants encountered first when using the Minpaku Guide, displayed multiple artifacts in a grid layout. To maximize heterogeneity, we selected images from a larger set using an agglomerative clustering method ($K = 50$) applied to image features. The features were extracted with ResNet50 [23], a deep learning convolutional neural network
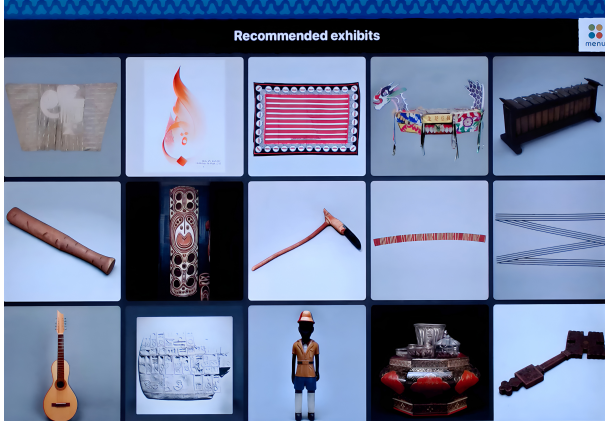
(a) Example of image grid page
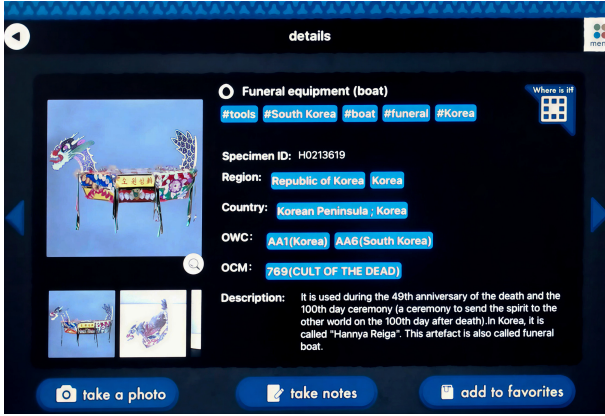
(b) Example of description page

Figure 1: Overview of the application interfaces

widely used in image processing research. The items displayed in the Minpaku Guide were identical for all participants.

3.1.2 Participants. Twenty-five participants took part in the experiment, representing different nationalities and ranging in age from 18 to 34 years. Participants were recruited using a snowball sampling strategy. Participants were informed about the purpose and procedures of the study, including its expected duration, potential risks, and policies regarding data storage and usage to safeguard privacy. Participants were also explicitly told that they could withdraw from the study at any point without penalty. All participants provided written informed consent prior to the experiment.

3.1.3 Apparatus. Eye tracking was conducted using a Tobii Pro Nano at a sampling rate of 60 Hz. The device, connected to a Dell Precision 7550 computer, was positioned above an iPad Pro 2022 to avoid obstructing participants during interaction. Each participant sat approximately 50 cm from the iPad.

Because a video capture card was unavailable, the screen activity of the iPad application, along with participants' gaze trajectories, was recorded using a Logitech C920 webcam. The webcam feed was shown on a separate monitor that was not visible to the participant.

Algorithm 1 Training Missing-Data-Aware Network for Personality Prediction

Require: Raw eye-tracking sequence $\{(g_t^x, g_t^y, p_t)\}_{t=1}^T$, sampling period Δt, window length L, number of epochs E
Ensure: Trained BiLSTM parameters $\theta = \{W, b\}$
1: Compute features:
   - Compute gaze velocity $v_t = \|\mathbf{g}_t - \mathbf{g}_{t-1}\|_2 / \Delta t$ ($v_1 = 0$).
   - Normalize gaze coordinates: $\bar{g}_t^x, \bar{g}_t^y \in [-1, 1]$.
   - Standardize pupil diameter p_t and velocity v_t with z-scores.
2: Handle missingness:
   - Replace NaN with 0.
   - Construct binary mask $\mathbf{m}_t \in \{0, 1\}^d$.
   - Update temporal gaps Δt_t recursively.
3: Build augmented input $\mathbf{f}_t = [\tilde{\mathbf{x}}_t, \mathbf{m}_t, \Delta t_t] \in \mathbb{R}^k$.
4: Partition sequence into overlapping windows $\mathbf{W}_k^{(n)} \in \mathbb{R}^{L \times k}$.
5: **for** epoch = 1 to E **do**
6:   **for** each window $\mathbf{W}_k^{(n)}$ with label $y^{(n)}$ **do**
7:     Encode window with BiLSTM $\rightarrow \{\mathbf{h}_t\}_{t=1}^L$.
8:     Summarize: $\mathbf{z} = [\overrightarrow{\mathbf{h}}_L; \overleftarrow{\mathbf{h}}_1]$.
9:     Predict with softmax: $\hat{y} = \text{softmax}(W\mathbf{z} + b)$.
10:     Compute loss: $\mathcal{L} = -\sum_{c=1}^C \mathbf{1}[y = c] \log \hat{y}_c$.
11:     Update parameters θ with gradient descent.
12:   **end for**
13: **end for**

Tracking was performed for both eyes and their recorded data was used for analysis. Recordings had a resolution of 1024 × 576 pixels with an average accuracy of 0.52°. Ambient lighting was kept consistent across participants to avoid effects on pupil dilation.

3.2 Missing-Data-Aware Network for Personality Prediction

We propose a missing-data-aware framework for predicting personality traits from eye-tracking signals. First, raw gaze coordinates, pupil size, and gaze velocity are represented as multivariate time series segmented into fixed-length windows (Section 3.2.1). Second, missingness is modeled with binary masks and temporal gap features, treating absent data as informative signals rather than noise (Section 3.2.3). Finally, the augmented sequences are processed by a bidirectional LSTM, which captures temporal dependencies and produces window-level predictions (Section 3.2.3). Together, these components form a pipeline that leverages both eye-movement patterns and missing data for personality prediction. An overview of the full pipeline is presented in Algorithm 1.

3.2.1 Eye-Tracking Data as a Multivariate Time Series. Let f_s denote the sampling frequency (Hz), and $\Delta t = 1/f_s$ the sampling period. That is, the eye tracker records f_s samples per second, with Δt seconds between two consecutive recordings.

For a recording session of length T samples, we represent the eye-tracking stream as a multivariate time series, where each sample is represented as follows:

$$\mathbf{x}_t = \begin{bmatrix} g_t^x, & g_t^y, & p_t, & v_t \end{bmatrix}^\top \in \mathbb{R}^d, \quad t = 1, \dots, T, \qquad d = 4,$$

where g_t^x, g_t^y are the horizontal and vertical gaze coordinates on the display, p_t is pupil diameter, and v_t is gaze velocity. Each $\mathbf{x}_t$ is therefore a four-dimensional snapshot of eye behavior at time t.

The gaze velocity is computed from the gaze position vector $\mathbf{g}_t = (g_t^x, g_t^y)^\top$ as

$$v_t = \frac{\|\mathbf{g}_t - \mathbf{g}_{t-1}\|_2}{\Delta t}, \quad t \geq 2,$$

with $v_1 = 0$ by convention. This captures how far the eyes moved between consecutive samples, divided by the elapsed time.

Session and window representation. We collect the full session as

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T] \in \mathbb{R}^{d \times T},$$

with timestamps $\tau_t = (t-1)\Delta t$. Here, $\mathbf{X}$ is arranged so that each column corresponds to one moment in time ($t = 1, \dots, T$), and each row corresponds to one of the $d = 4$ recorded features (horizontal gaze g^x, vertical gaze g^y, pupil size p, velocity v). Thus, $\mathbf{X}$ can be seen as a compact table of the entire session: moving across columns follows the sequence of time steps, while moving down rows inspects the different measurements collected at that time.

For model training, this long sequence is divided into overlapping subsequences (windows) of fixed length L:

$$\mathbf{W}_k = [\mathbf{x}_{s_k}, \mathbf{x}_{s_k+1}, \dots, \mathbf{x}_{s_k+L-1}] \in \mathbb{R}^{d \times L}, \quad k = 1, \dots, K,$$

where s_k is the start index of the k-th window and K is the total number of extracted windows. Each $\mathbf{W}_k$ is therefore a short clip of the session that spans L consecutive timesteps but keeps all d features at each step. In practice, this sliding-window procedure allows the model to learn from many local fragments of behavior, capturing recurring gaze patterns that may be predictive of personality traits.

Normalization. Let (W, H) denote the screen width and height in pixels, and (g_t^x, g_t^y) the raw gaze position recorded at time t in pixel coordinates. To remove dependence on the specific display size, gaze coordinates are rescaled to a zero-centered, unit-square system: $\bar{g}_t^x = \frac{2g_t^x}{W} - 1$, $\bar{g}_t^y = \frac{2g_t^y}{H} - 1$, so that $(\bar{g}_t^x, \bar{g}_t^y) \in [-1, 1]^2$. After this transformation, the center of the screen corresponds to $(0, 0)$, the left and right edges correspond to -1 and $+1$ on the x-axis, and the top and bottom edges correspond to $+1$ and -1 on the y-axis. This mapping ensures that gaze positions are expressed in a common reference frame across devices and participants.

For the pupil diameter, let p_t be the raw measurement at time t, μ_p the mean pupil diameter across the training set, and σ_p the corresponding standard deviation. We standardize pupil diameter using z-score normalization: $\tilde{p}_t = \frac{p_t - \mu_p}{\sigma_p}$.

Similarly, let v_t denote the raw gaze velocity, with μ_v and σ_v the mean and standard deviation of velocity over the training set. Velocity is standardized in the same way: $\tilde{v}_t = \frac{v_t - \mu_v}{\sigma_v}$.

This standardization rescales p_t and v_t to have zero mean and unit variance, so that the model focuses on relative fluctuations (e.g., dilations above or below a typical pupil size, or faster versus slower gaze shifts) rather than absolute raw values, which may vary considerably across individuals.

Final feature vector. Unless otherwise noted, the input at each timestep t is the normalized feature vector

$$\tilde{\mathbf{x}}_t = [\bar{g}_t^x, \bar{g}_t^y, \tilde{p}_t, \tilde{v}_t]^\top \in \mathbb{R}^4,$$

which stacks together the four processed measurements: normalized horizontal gaze position $\bar{g}_t^x$, normalized vertical gaze position $\bar{g}_t^y$, standardized pupil diameter $\tilde{p}_t$, and standardized gaze velocity $\tilde{v}_t$.

3.2.2 Modeling Missingness in Eye-Tracking Data. Eye-tracking data collected under naturalistic conditions often contain substantial missing values. This missingness arises when the participant blinks, looks away from the screen, or when the tracker momentarily loses calibration. In the raw data, such missing entries are recorded as NaN. Since the model cannot operate directly on NaN values, we replace them with zeros before training. However, the value 0 can also be a valid observation (for example, a gaze coordinate of 0 corresponds to the center of the screen). To prevent confusion between true zeros and placeholders for missing values, we introduce an explicit binary validity mask.

Binary validity masks. For each feature dimension $j \in \{1, \dots, d\}$ at time t, we define a binary indicator variable

$$m_t^j = \begin{cases} 1, & \text{if feature } j \text{ has a valid observed value at time } t, \\ 0, & \text{if feature } j \text{ is missing (recorded as NaN in the raw data),} \end{cases}$$

where $d = 4$ in our case (horizontal gaze, vertical gaze, pupil size, and velocity). These indicators are then collected into a mask vector

$$\mathbf{m}_t = [m_t^1, m_t^2, \dots, m_t^d]^\top \in \{0, 1\}^d.$$

The role of this mask is to preserve the distinction between two different situations: (1) a feature truly takes the value 0 (e.g., $\bar{g}_t^x = 0$ meaning the gaze is exactly at the horizontal center of the screen), and (2) the original data at that position was missing and has been replaced with 0 only as a placeholder so the model can process the input. Without the mask, these two cases would be indistinguishable.

Concretely, if the horizontal gaze coordinate is missing at time t, we set its value in $\tilde{\mathbf{x}}_t$ to 0 but also set $m_t^1 = 0$. If the gaze is genuinely at the screen center, then the value is also 0 but the mask records $m_t^1 = 1$. Thus, the pair $(\tilde{\mathbf{x}}_t, \mathbf{m}_t)$ allows the model to differentiate between a true zero measurement and a zero that only indicates missingness.

Temporal gap encoding. In addition to the binary masks, we record how long each feature has been continuously missing. For feature j at time t, we define a temporal gap variable Δt_t^j that is updated recursively as

$$\Delta t_t^j = \begin{cases} 0, & \text{if } m_t^j = 1 \text{ (feature } j \text{ is observed at time } t), \\ \Delta t_{t-1}^j + \Delta t, & \text{if } m_t^j = 0 \text{ (feature } j \text{ is missing at time } t), \end{cases}$$

where Δt is the sampling period.

Thus, Δt_t^j counts how long feature j has been missing up to time t. Whenever a valid measurement is observed ($m_t^j = 1$), the gap resets to 0. When the feature remains missing across consecutive timesteps ($m_t^j = 0$), the gap grows by Δt each time step.

This variable provides temporal context for the missingness:

- If Δt_t^j is small (close to 0), the absence is likely due to a short interruption such as a blink or a brief calibration error.
- If Δt_t^j grows large, it indicates a prolonged disengagement, for example the participant looking away from the screen for several seconds.

By including Δt_t^j as an explicit feature, the model can distinguish between short, transient dropouts and longer episodes of missing data, which may carry different behavioral meanings.

Extended feature representation. At each timestep t, we construct an augmented input vector

$$\mathbf{f}_t = \begin{bmatrix} \tilde{\mathbf{x}}_t, & \mathbf{m}_t, & \Delta t_t \end{bmatrix},$$

where:

- $\tilde{\mathbf{x}}_t \in \mathbb{R}^d$ are the normalized feature values (horizontal and vertical gaze coordinates, standardized pupil size, and standardized velocity),
- $\mathbf{m}_t \in \{0, 1\}^d$ is the binary validity mask, indicating for each feature whether the value at time t was truly observed (1) or was originally missing and replaced by a placeholder (0),
- $\Delta t_t = [\Delta t_t^1, \ldots, \Delta t_t^d]^\top \in \mathbb{R}^d$ contains the temporal gap variables, which record how long each feature has been continuously missing.

This yields an augmented representation of dimension $k = 3d$: for each of the d base features, we include its normalized value, a validity flag, and a temporal gap duration.

The motivation for this augmentation is threefold:

(1) The model can *ignore invalid entries* by using the mask $\mathbf{m}_t$ while still retaining information about which features were missing.
(2) The temporal gap variables Δt_t allow the model to capture behavioral patterns such as the difference between a short blink (a brief gap) and sustained disengagement (a long gap).
(3) By combining observed values with missingness structure, the model can potentially learn *personality-related regularities*, for example, that certain individuals tend to look away more often or for longer periods, which may be predictive of traits like neuroticism or conscientiousness.

3.2.3 *Sequential Modeling for Personality Prediction.* Personality prediction from eye-tracking is framed as a sequence-to-label task: the model receives a short sequence of eye-tracking data and must predict the participant's personality traits.

Given an augmented feature window

$$\mathbf{W}_k^{(n)} = [\mathbf{f}_{s_k}, \mathbf{f}_{s_k+1}, \ldots, \mathbf{f}_{s_k+L-1}]^\top \in \mathbb{R}^{L \times k},$$

we stack L consecutive augmented vectors $\mathbf{f}_t \in \mathbb{R}^k$ (normalized features, validity masks, and temporal gaps). Here, n indexes the participant, s_k is the start index of the k-th window, and k is the feature dimension.

The prediction target is $y^{(n)} \in \{1, 2, 3\}^5$, a 5-dimensional vector for the Big Five traits (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism), where each entry takes values 1, 2, 3 for *Low*, *Medium*, and *High*. For the classification setup, participants' continuous personality scores (originally measured on a 1–5 scale) were partitioned into three groups using a quantile-based procedure following [61]. For each trait, cut points were set at the 33rd and 66th percentiles of its empirical distribution, so that each group contained roughly the same number of participants. This procedure yielded balanced class sizes and provided clear separation among low, medium, and high scorers. All windows from participant n share the same label $y^{(n)}$, so the task is to learn a mapping from temporal sequences $\mathbf{W}_k^{(n)}$ to the trait categories.

Sequential representation learning. Prior work often summarized windows into handcrafted statistics, such as average fixation duration, blink counts, or velocity variance. While useful, such summary features discard the temporal ordering of the signal. Here we instead retain the sequence itself, enabling the model to exploit fine-grained temporal dependencies—such as sustained fixations, repeated scanning motions, or short pupil dilations—that may be characteristic of personality traits.

BiLSTM encoder. To capture temporal dependencies in the data, we employ a bidirectional Long Short-Term Memory (BiLSTM) network. An LSTM is a type of recurrent neural network that is designed to process sequences one element at a time, while retaining information from previous steps through a hidden state. This makes it well suited to modeling time-series signals such as eye movements, where the current behavior depends strongly on what came before.

At each timestep t, the hidden state is updated as

$$\mathbf{h}_t = \mathrm{BiLSTM}(\mathbf{f}_t, \mathbf{h}_{t-1}), \quad \mathbf{h}_t \in \mathbb{R}^h,$$

where $\mathbf{f}_t$ is the augmented feature vector at time t and h is the dimensionality of the hidden state.

Unlike a standard LSTM, which only processes the sequence forward in time, a BiLSTM maintains two parallel chains of hidden states: a forward chain $\overrightarrow{\mathbf{h}}_t$ that processes the sequence from $1 \rightarrow L$, and a backward chain $\overleftarrow{\mathbf{h}}_t$ that processes it in reverse from $L \rightarrow 1$. At each timestep, the two are concatenated as $\mathbf{h}_t = \begin{bmatrix} \overrightarrow{\mathbf{h}}_t; & \overleftarrow{\mathbf{h}}_t \end{bmatrix}$.

This bidirectional setup is important because the meaning of an event often depends on its temporal context. For example, a brief period of missing data could be interpreted as a blink if it is followed immediately by normal gaze behavior, but the same gap might indicate disengagement if it occurs before and after long stretches of missingness. By looking both backward and forward in time, the BiLSTM can capture such context more effectively than a unidirectional model.

Window-level representation and prediction. After processing a window of L timesteps, the BiLSTM produces a sequence of hidden states $\{\mathbf{h}_t\}_{t=1}^L$, each encoding information about the input at time t and its surrounding context. To obtain a fixed-length representation for the entire window, we concatenate the last forward hidden state $\overrightarrow{\mathbf{h}}_L$ (which summarizes information up to the end of the window) with the last backward hidden state $\overleftarrow{\mathbf{h}}_1$ (which summarizes information looking backward from the start): $\mathbf{z} = \begin{bmatrix} \overrightarrow{\mathbf{h}}_L; & \overleftarrow{\mathbf{h}}_1 \end{bmatrix}$.

This vector $\mathbf{z}$ serves as a compact summary of the whole sequence, capturing both past and future dependencies.

The representation $\mathbf{z}$ is passed to a classification head, which applies a linear transformation followed by a softmax activation, outputting a probability distribution over the three categories for each personality trait.

Training objective. Model training minimizes the categorical cross-entropy loss: $\mathcal{L} = -\sum_{c=1}^{C} \mathbf{1}[y = c] \log \hat{y}_c$,

which penalizes the divergence between the predicted distribution $\hat{y}$ and the true label y. Here, $\mathbf{1}[y = c]$ is an indicator function that equals 1 when the ground-truth label is c and 0 otherwise. Intuitively, this loss encourages the model to assign high probability to the correct class while discouraging probability mass on incorrect categories.

4 Experimental Evaluation

4.1 Classifier Training

We train a missing-data-aware *bidirectional* LSTM classifier using a 12-dimensional input feature vector at each timestep: (1) horizontal gaze coordinate g_x, (2) its validity mask m_{g_x}, (3) temporal gap Δt_{g_x}, (4) vertical gaze coordinate g_y, (5) mask m_{g_y}, (6) gap Δt_{g_y}, (7) gaze velocity v, (8) mask m_v, (9) gap Δt_v, (10) pupil diameter p, (11) mask m_p, and (12) gap Δt_p. Sequences are segmented with a sliding window of length L=100 and overlap = 50 (stride = 50) [64], which corresponds to a window duration of 1.67 s with 0.83 s stride at a 60 Hz sampling rate. The network uses hidden size 64, 2 layers, dropout 0.3, and bidirectionality, followed by a two-layer classifier head; weights are initialized with orthogonal (LSTM) and Xavier (linear) schemes. Training uses cross-entropy loss with Adam optimization (learning rate 10^{-3}, weight decay 10^{-5}), gradient clipping, and a ReduceLROnPlateau scheduler (patience = 10, factor = 0.5), for up to 100 epochs with early stopping after 15 epochs without validation improvement. We evaluate performance using stratified 5-fold cross-validation (preserving the Low/Medium/High class distribution) and report mean accuracy and macro-averaged F1 scores with standard deviations. To prevent data leakage, we adopt a leakage-free data-splitting strategy in which data are split at the segment level: for each participant, continuous signals are first divided into non-overlapping contiguous segments, which are then assigned to one of the five folds. Sliding windows are generated only within segments after fold assignment, ensuring no overlap in raw data or windows between training and test folds. We also conduct participant-stratified cross-validation, assigning each participant entirely to one fold to test generalizability to unseen individuals, and report average macro F1 with standard deviations. Leave-one-subject-out (LOSO) was not feasible because each participant only provided data for one class. As a result, test folds contained a single class only, making accuracy and F1 computation infeasible.

4.2 Results

Table 1 summarizes the classification performance for each of the Big Five traits. Overall, the proposed missing-data-aware BiLSTM achieves stable and moderately strong performance across all traits, with mean accuracies ranging from 73.09% (Agreeableness) to 77.69% (Neuroticism). Neuroticism (77.69%) and Openness (77.52%) exhibited relatively higher predictive performance. Extraversion (76.69%) and Conscientiousness (74.52%) also achieved competitive results. Agreeableness remains the most challenging trait, with macro F1 score at 73.09%. Figure 2 depicts the confusion matrices for each of the Big Five traits.

Table 1: Classification performance (mean ± standard deviation) across five folds for each Big Five trait using the full pipeline (time series + masks + temporal gaps). The last column reports participant-stratified cross-validation (by-participant) macro F1 averaged across folds. Values are percentages; the % sign is omitted for readability.

Trait	Accuracy	Macro F1	By-Participant F1
Openness	77.96 ± 2.42	77.52 ± 2.63	66.23 ± 7.92
Conscientiousness	74.86 ± 2.34	74.52 ± 2.32	65.99 ± 25.12
Extraversion	78.17 ± 1.49	76.69 ± 1.24	70.89 ± 13.92
Agreeableness	73.76 ± 1.86	73.09 ± 2.05	63.13 ± 15.20
Neuroticism	79.19 ± 2.70	77.69 ± 2.92	74.33 ± 23.85

In addition to fold-level cross-validation, we evaluated performance using participant-stratified cross-validation to test generalizability to unseen individuals. While performance naturally decreased in this stricter setting (e.g., Macro F1 score of 63.1% for Agreeableness and 66.2% for Openness), the model still achieved reasonable performance, with Extraversion (70.9%) and Neuroticism (74.3%) showing the strongest generalization.

4.3 Ablation Study

To examine whether incorporating missingness information affects the model performance, we conducted a set of ablation experiments where different subsets of features and modeling strategies were retained. Here, binary masks denote missingness indicators (1 = valid, 0 = missing).

(1) **Full Pipeline.** The proposed missing-data-aware BiLSTM trained with the complete representation: time-series features (horizontal/vertical gaze, pupil diameter, gaze velocity), their binary validity masks, and temporal gap encodings.

(2) **Time Series + Temporal Gap.** The BiLSTM is trained using the four time-series features together with temporal gap (Δt) encodings, but without missingness indicators. In this setting, the model retains information about the elapsed time since the last observation, but does not explicitly encode whether a value is observed or missing.

(3) **Time Series Only.** The BiLSTM trained solely on the four raw time-series features, without masks or temporal gaps. This serves as a baseline for sequential modeling without any explicit missingness indicators.

(4) **Non-sequential Baseline.** A non-sequential model trained on handcrafted features. For each of the four base signals (gaze x, gaze y, pupil, gaze velocity), we compute five descriptive statistics (minimum, maximum, mean, standard deviation, median), yielding a 20-dimensional feature vector. A random forest classifier is then trained on this representation, providing a feature-engineered baseline without temporal modeling.

Table 2 summarizes the results of the ablation study across the Big Five traits. The full pipeline, which combines time-series features, explicit missingness masks, and temporal gap encodings, consistently achieves the best performance across traits, with accuracies ranging from approximately 74% (Agreeableness) to 79% (Neuroticism) and macro F1 scores between 73% and 78%. Removing missingness masks while retaining temporal gap encodings

Table 2: Ablation study results across Big Five traits. "Full" uses time series + masks + temporal gaps features; "TS+Temporal Gap" uses time series + temporal gap features; "TS Only" uses raw time series features; and "Statistical" uses handcrafted statistical features with a Random Forest. Results are reported as mean accuracy and macro F1 (%) with standard deviation.

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
	Accuracy (%)				
Full	**77.96 ± 2.42**	**74.86 ± 2.34**	**78.17 ± 1.49**	**73.76 ± 1.86**	**79.19 ± 2.70**
TS+Temporal Gap	69.96 ± 1.50	70.43 ± 1.48	73.49 ± 2.65	67.15 ± 1.70	71.80 ± 0.99
TS Only	69.23 ± 2.48	66.86 ± 0.74	72.67 ± 2.32	65.75 ± 1.36	73.44 ± 0.44
Statistical	67.59 ± 1.84	67.15 ± 2.18	70.97 ± 3.51	65.56 ± 1.67	73.73 ± 2.08
	Macro F1 Score (%)				
Full	**77.52 ± 2.63**	**74.52 ± 2.32**	**76.69 ± 1.24**	**73.09 ± 2.05**	**77.69 ± 2.92**
TS+Temporal Gap	69.87 ± 1.54	70.38 ± 1.52	70.29 ± 3.12	66.21 ± 1.68	69.60 ± 2.58
TS Only	68.73 ± 2.45	66.51 ± 0.71	70.64 ± 2.12	65.56 ± 1.46	71.56 ± 0.37
Statistical	66.40 ± 1.80	66.99 ± 2.27	67.85 ± 3.30	64.62 ± 1.57	70.31 ± 2.90

(*TS+Temporal Gap*) leads to consistent performance degradation across all traits, suggesting that explicit missingness indicators provide complementary information beyond temporal gap encodings alone. Using only raw time-series features (*TS Only*) further reduces performance, indicating that modeling both irregular sampling and missing data is important for effective temporal representation learning. The statistical baseline, which relies on handcrafted features and a Random Forest classifier, performs worst overall, highlighting the benefit of sequential models augmented with explicit temporal and missingness representations. Overall, the ablation results demonstrate that each component of the proposed pipeline contributes meaningfully to robust personality trait prediction.

5 Discussion

The results in Table 1 show that it is indeed possible to classify personality traits from temporal behavioral features, and that this approach works best when the full modeling pipeline is used – that is, when raw time-series signals are combined with missingness and temporal gap information. The consistently strong scores suggest that eye-tracking data contains rich signals about stable psychological traits. At the same time, the differences in performance across the Big Five traits reveal how personality is expressed unevenly in gaze behavior. These patterns provide not only theoretical insights into the behavioral expression of personality but also practical guidance for how to build computational models.

Among the five traits, *Neuroticism (N)* and *Openness (O)* showed relatively stronger and more consistent performance, in line with prior work linking these traits to distinctive gaze variability and exploratory viewing behaviors [1, 57, 59, 60]. *Extraversion (E)* and *Conscientiousness (C)* achieved moderate performance, suggesting that attentional cues are informative but less distinctive in solitary tasks [29, 43, 69]. In contrast, *Agreeableness (A)* remained the most difficult trait to predict, consistent with evidence that it is primarily expressed in social interaction contexts [59, 73]. Overall, traits tied to internal regulation (attention, emotion) were easier to detect from gaze than socially oriented traits, which may require richer contexts or additional modalities.

When examining generalization, the by-participant cross-validation revealed a different picture. Scores were lower and more variable

across individuals, especially for Agreeableness and Conscientiousness. This highlights a major challenge: while group-level models capture broad tendencies, predicting personality consistently across different individuals is more complex. One reason is the heterogeneity of cognitive strategies: individuals may differ in how they allocate attention, process information, or regulate effort during a task. For example, some people naturally adopt systematic scanning routines, while others rely on more opportunistic or intuitive search behaviors [6, 29], leading to divergent gaze dynamics even if they share the same personality trait [48]. Another possible explanation is the limited sample size: with only 25 participants, the distribution of Big Five personality traits may differ substantially from that observed in much larger populations [67]. This discrepancy can influence performance at the individual-participant level and limit the model's ability to generalize. Increasing the number and diversity of participants would likely lead to a more representative trait distribution and improve generalization performance.

Behavioral habits and situational factors further complicate generalization. Gaze patterns are shaped not only by personality but also by momentary states such as fatigue, stress, or task engagement [19, 75]. A conscientious individual (Conscientiousness) might display high attentional stability in one session but exhibit more variability under distraction or cognitive load. Similarly, Agreeableness – already subtle in solitary tasks – may only surface in socially interactive contexts, making it less consistently observable across participants [73].

These considerations suggest that variability is not simply "noise" but reflects meaningful person–context interactions. Capturing these nuances may require models that account for both stable dispositions and dynamic behavioral states. Approaches such as domain adaptation, personalized modeling, or hierarchical frameworks could help bridge the gap, enabling models to distinguish between trait-driven regularities and state-driven fluctuations. More broadly, these findings underscore the importance of integrating cognitive and behavioral theories into computational modeling: personality is not expressed uniformly, but filtered through the individual's strategies, habits, and situational context [19].

The ablation results in Table 2 emphasize why temporal context matters. The complete pipeline (time series + masks + temporal
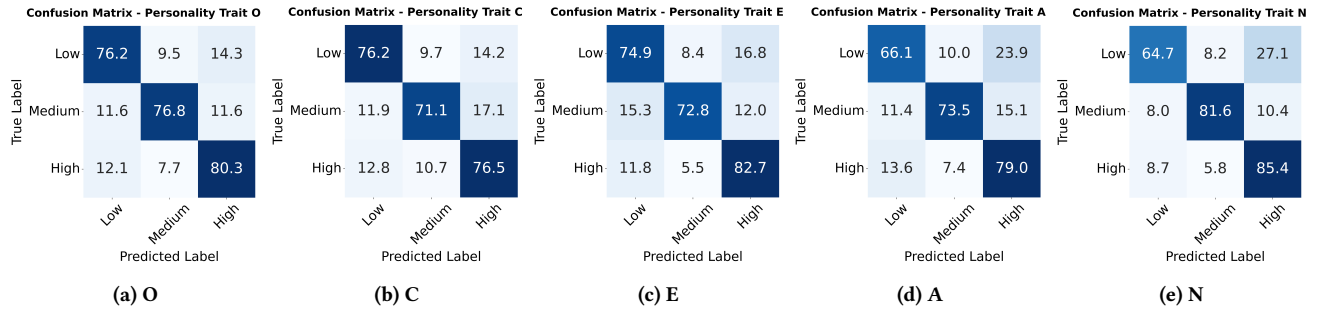
Figure 2: Confusion matrices for each personality trait.

gaps) consistently outperformed both raw time series alone and handcrafted statistical features. Notably, the inclusion of masks and temporal gap information added substantial value. This suggests that personality-relevant information is not just in the sequence of gaze coordinates but also in irregularities – when people look away, how long gaps last, and how missing data is structured. This aligns with recent research in time-series modeling that treats missingness as informative rather than as noise [11, 46]. Ignoring such irregularities risks discarding psychologically meaningful signals.

From a cognitive perspective, irregularities reflect the underlying processes of attention and information processing. Long gaps can signal lapses in sustained attention or shifts in cognitive focus, often linked to high Neuroticism or lower engagement with the task [16]. Doherty-Sneddon and Phelps [16] noted that people tend to look away when under high cognitive load. By contrast, more frequent gaze shifts and less tightly constrained fixation patterns may reflect exploratory viewing strategies, which are characteristic of individuals high in Openness, who tend to engage flexibly with stimuli and seek novel information. Thus, the timing of gaze interruptions offers indirect but meaningful cues about attention and control.

Behaviorally, missingness patterns capture differences in task approach. Some individuals show bursts of exploratory scanning interspersed with pauses – consistent with curiosity and Openness – while others show repetitive fixations and fewer breaks, reflecting more rigid or habitual engagement [1]. This means irregular timing is not just a random error but a behavioral marker of personality-linked cognitive styles.

The methodological implications are equally important. Traditional feature engineering often assumes that missing data should be interpolated or discarded, but our findings indicate that the very presence and distribution of missingness can serve as predictive features. This perspective aligns with a growing body of behavioral informatics research, which argues that "what is absent" in behavioral traces can be as informative as "what is present" [11]. For personality modeling, this means that gaze dynamics should be analyzed not only for their visible sequences but also for their absences, interruptions, and irregular rhythms.

In summary, our findings show that eye-tracking holds promise for inferring personality, albeit with important boundaries. Traits tied to attention and emotion, such as Openness and Neuroticism, are especially well captured, while socially expressed trait like Agreeableness may require richer data sources such as language, physiology, or social interaction [68, 69]. Task design also plays a

critical role: social or collaborative contexts may amplify expressions of traits that remain muted in solitary tasks.

In general, the temporal structure of eye movements, including where gaps occur, provides meaningful cues about personality. However, not all traits are equally predictable, and performance varies across individuals. Recognizing these nuances is crucial both for advancing personality science and for building adaptive systems that personalize interactions. Combining eye-tracking with multimodal data and more diverse task designs could lead to models that are both more comprehensive and more generalizable.

6 Conclusion

This study demonstrates that eye-tracking data, especially when analyzed as time-based sequences that include gaps and missing points, can provide useful insights into personality. We found that traits related to attention and emotion regulation, such as Neuroticism and Openness, are captured well in gaze patterns. In contrast, socially driven traits like Agreeableness is harder to detect in tasks performed alone. This highlights both the promise and the limits of using eye movements alone to infer personality.

Beyond prediction accuracy, our work makes two broader contributions. First, we demonstrate that irregularities in gaze, moments when people look away or pause, are not just noise but carry meaningful signals about attention control, disengagement, and exploration. Second, we highlight that personality is shaped by both stable traits and changing contexts, which means models must account for individual differences while also adapting to situational influences.

Limitations. Our study focused exclusively on museum settings using picture-based search systems. Additionally, different eye-tracking devices may yield varying results due to differences in processing capabilities and other influencing factors, such as sampling frequency, camera resolution, and calibration quality.

Future Work. Future research could extend this work to a wider range of interactive settings. Additionally, combining eye-tracking data with other modalities such as electrodermal activity (EDA), electroencephalography (EEG), or electrocardiography (ECG) could enrich multimodal models and offer deeper insights into how personality traits influence information-seeking behavior.

Acknowledgments

References

[1] Sergio Agnoli, Laura Franchin, Enrico Rubaltelli, and Giovanni Emanuele Corazza. 2015. An eye-tracking analysis of irrelevance processing as moderator of openness and creative performance. *Creativity Research Journal* 27, 2 (2015), 125–132.

[2] Hosam Al-Samarraie, Atef Eldenfria, and Husameddin Dawoud. 2017. The impact of personality traits on users' information-seeking behavior. *Information Processing & Management* 53, 1 (2017), 237–247.

[3] Brooke J Arterberry, Matthew P Martens, Jennifer M Cadigan, and David Rohrer. 2014. Application of generalizability theory to the big five inventory. *Personality and individual differences* 69 (2014), 98–103.

[4] Neal Ashkanasy, Paul L Bowen, Fiona H Rohde, and Chiu Yueh Alice Wu. 2007. The effects of user characteristics on query performance in the presence of information request ambiguity. *Journal of Information Systems* 21, 1 (2007), 53–82.

[5] Nicholas J Belkin. 1990. The cognitive viewpoint in information science. *Journal of information science* 16, 1 (1990), 11–15.

[6] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. 2019. Detecting personality traits using eye-tracking data. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[7] Hal Blumenfeld. 2002. Neuroanatomy Through Clinical Cases.

[8] Nattapat Boonprakong, Benjamin Tag, Jorge Goncalves, and Tilman Dingler. 2025. How Do HCI Researchers Study Cognitive Biases? A Scoping Review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[9] Iain Campbell and Keith Van Rijsbergen. 1996. The ostensive model of developing information needs. In *CoLIS 2: second international conference on conceptions of library science: integration in perspective (Copenhague, October 13-16, 1996)*. 251–268.

[10] R Nicholas Carleton, MA Peter J Norton, and Gordon JG Asmundson. 2007. Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of anxiety disorders* 21, 1 (2007), 105–117.

[11] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. 2018. Recurrent neural networks for multivariate time series with missing values. *Scientific reports* 8, 1 (2018), 6085.

[12] Li Chen, Wanling Cai, Dongning Yan, and Shlomo Berkovsky. 2023. Eye-tracking-based personality prediction with recommendation interfaces. *User Modeling and User-Adapted Interaction* 33, 1 (2023), 121–157.

[13] Michael J Cole, Jacek Gwizdka, Chang Liu, Nicholas J Belkin, and Xiangmin Zhang. 2013. Inferring User Knowledge Level from Eye Movement Patterns. *Information Processing & Management* 49, 5 (2013), 1075–1091.

[14] Michael J Cole, Chathra Hendahewa, Nicholas J Belkin, and Chirag Shah. 2015. User activity patterns during information search. *ACM Transactions on Information Systems (TOIS)* 33, 1 (2015), 1–39.

[15] Sahraoui Dhelim, Nyothiri Aung, Mohammed Amine Bouras, Huansheng Ning, and Erik Cambria. 2022. A survey on personality-aware recommendation systems. *Artificial Intelligence Review* 55, 3 (2022), 2409–2454.

[16] Gwyneth Doherty-Sneddon and Fiona G Phelps. 2005. Gaze aversion: A response to cognitive or social difficulty? *Memory & cognition* 33, 4 (2005), 727–733.

[17] EM Donahue, OP John, and RL Kentle. 1991. The Big Five Inventory-Versions 4a and 54, Berkeley. *CA: University of California, Berkeley, Institute of Personality and Social Research* (1991).

[18] Carsten Eickhoff, Sebastian Dungs, and Vu Tran. 2015. An Eye-Tracking Study of Query Reformulation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, 13–22.

[19] William Fleeson and Patrick Gallagher. 2009. The implications of Big Five standing for the distribution of trait manifestation in behavior: fifteen experience-sampling studies and a meta-analysis. *Journal of personality and social psychology* 97, 6 (2009), 1097.

[20] Léon Franzen, Amanda Cabugao, Bianca Grohmann, Karine Elalouf, and Aaron P Johnson. 2022. Individual pupil size changes as a robust indicator of cognitive familiarity differences. *PloS one* 17, 1 (2022), 1–22.

[21] Jiaman He, Zikang Leng, Dana McKay, Damiano Spina, and Johanne R. Trippas. 2025. Can We Hide Machines in the Crowd? Quantifying Equivalence in LLM-in-the-loop Annotation Tasks. In *Proceedings of the 2025 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*. 426–436.

[22] Jiaman He, Zikang Leng, Dana McKay, Johanne R Trippas, and Damiano Spina. 2025. Characterising Topic Familiarity and Query Specificity Using Eye-Tracking Data. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2602–2606.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[24] Jannica Heinström. 2003. Five personality dimensions and their influence on information behaviour. *Information research* 9, 1 (2003), 9–1.

[25] Shuk Ying Ho, Michael J Davern, and Kar Yan Tam. 2008. Personalization and choice behavior: the role of personality traits. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* 39, 4 (2008), 31–47.

[26] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[27] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures.* oup Oxford.

[28] Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao, and Jimeng Sun. 2020. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in biology and medicine* 122 (2020), 103801.

[29] Sabrina Hoppe, Tobias Loetscher, Stephanie A Morey, and Andreas Bulling. 2018. Eye movements during everyday behavior predict personality traits. *Frontiers in human neuroscience* 12 (2018), 328195.

[30] Rong Hu and Pearl Pu. 2011. Enhancing collaborative filtering systems with personality information. In *Proceedings of the fifth ACM conference on Recommender systems*. 197–204.

[31] Apple Inc. 2025. Apple Vision Pro. https://www.apple.com/apple-vision-pro/ Accessed: 2025-01-17.

[32] Hayley K Jach, Colin G DeYoung, and Luke D Smillie. 2022. Why do people seek information? The role of personality traits and situation perception. *Journal of Experimental Psychology: General* 151, 4 (2022), 934.

[33] Hayley K Jach and Luke D Smillie. 2019. To fear or fly to the unknown: Tolerance for ambiguity and Big Five personality traits. *Journal of Research in Personality* 79 (2019), 67–78.

[34] Vahid Sadiri Javadi, Fryderyk Róg, Aska Aksa, Johanne R. Trippas, Svitlana Vakulenko, and Lucie Flek. 2026. CHARISMA: Character-Based Interaction Simulation with Multi-LLM Agents Toward Computational Social Psychology. In *Proceedings of the ACM Conference on Information Interaction and Retrieval (CHIIR'26)*. 1–5.

[35] Kaixin Ji, Danula Hettiachchi, Flora D. Salim, Falk Scholer, and Damiano Spina. 2024. Characterizing Information Seeking Processes with Multiple Physiological Signals. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. 1006–1017.

[36] Oliver P John, Sanjay Srivastava, et al. 1999. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. (1999).

[37] Min Jeong Kang, Ming Hsu, Ian M Krajbich, George Loewenstein, Samuel M McClure, Joseph Tao-yi Wang, and Colin F Camerer. 2009. The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological science* 20, 8 (2009), 963–973.

[38] Raghav Pavan Karumur, Tien T Nguyen, and Joseph A Konstan. 2018. Personality, user preferences and behavior in recommender systems. *Information Systems Frontiers* 20, 6 (2018), 1241–1265.

[39] Todd B Kashdan, Melissa C Stiksma, David J Disabato, Patrick E McKnight, John Bekier, Joel Kaji, and Rachel Lazarus. 2018. The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality* 73 (2018), 130–149.

[40] Jaikishan Khatri, Javier Marín-Morales, Masoud Moghaddasi, Jaime Guixeres, Irene Alice Chicchi Giglioli, and Mariano Alcañiz. 2022. Recognizing personality traits using consumer behavior patterns in a virtual retail store. *Frontiers in psychology* 13 (2022), 752073.

[41] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2176–2184.

[42] Martin Längkvist, Lars Karlsson, and Amy Loutfi. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern recognition letters* 42 (2014), 11–24.

[43] Thomas Le Bras, Benoit Allibe, and Karine Doré-Mazars. 2024. The way we look at an image or a webpage can reveal personality traits. *Scientific Reports* 14, 1 (2024), 15488.

[44] Zikang Leng, Megha Thukral, Yaqi Liu, Hrudhai Rajasekhar, Shruthi K. Hiremath, Jiaman He, and Thomas Plötz. 2025. AgentSense: Virtual Sensor Data Generation Using LLM Agents in Simulated Home Environments. arXiv:2506.11773 [cs.CV] https://arxiv.org/abs/2506.11773

[45] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting* 37, 4 (2021), 1748–1764.

[46] Zachary C Lipton, David C Kale, Randall Wetzel, et al. 2016. Modeling missing data in clinical time series with rnns. *Machine Learning for Healthcare* 56, 56 (2016), 253–270.

[47] Chenglong Ma, Ziqi Xu, Yongli Ren, Danula Hettiachchi, and Jeffrey Chan. 2025. PUB: An LLM-Enhanced Personality-Driven User Behaviour Simulator for Recommender System Evaluation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25).* 2690–2694.

[48] Andrew MacFarlane, George Buchanan, Areej Al-Wabil, Gennady Andrienko, and Natalia Andrienko. 2017. Visual analysis of dyslexia on search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval.* 285–288.

[49] Kristian E Markon. 2009. Hierarchies in the structure of personality traits. *Social and Personality Psychology Compass* 3, 5 (2009), 812–826.

[50] Sebastiaan Mathôt, Jasper Fabius, Elle Van Heusden, and Stefan Van der Stigchel. 2018. Safe and Sensible Preprocessing and Baseline Correction of Pupil-size Data. *Behavior research methods* 50 (2018), 94–106.

[51] Kenya Matsumoto, Sanae Shibata, Shimosato Seiji, Chizuru Mori, and Kunihiko Shioe. 2010. Factors influencing the processing of visual information from nonverbal communications. *Psychiatry and clinical neurosciences* 64, 3 (2010), 299–308.

[52] Robert R McCrae and Paul T Costa. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology* 52, 1 (1987), 81.

[53] Annika Melinder, Tim Brennen, Mikael Falkhaugen Husby, and Olav Vassend. 2020. Personality, confirmation bias, and forensic interviewing performance. *Applied Cognitive Psychology* 34, 5 (2020), 961–971.

[54] Martijn Millecamp, Cristina Conati, and Katrien Verbert. 2021. Classifeye: classification of personal characteristics based on eye tracking data in a recommender system interface. In *Joint Proceedings of the ACM IUI 2021 Workshops co-located with 26th ACM Conference on Intelligent User Interfaces (ACM IUI 2021)*, Vol. 2903.

[55] Melissa Onori, Alessandro Micarelli, and Giuseppe Sansonetti. 2016. A Comparative Analysis of Personality-Based Music Recommender Systems.. In *Empire@RecSys.* 55–59.

[56] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.

[57] Susan B Perlman, James P Morris, Brent C Vander Wyk, Steven R Green, Jaime L Doyle, and Kevin A Pelphrey. 2009. Individual differences in personality predict how people look at faces. *PloS one* 4, 6 (2009), e5952.

[58] Tobii Pro. 2025. Tobii Pro Glasses 3. https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3 Accessed: 2025-01-17.

[59] John F Rauthmann, Christian T Seubert, Pierre Sachse, and Marco R Furtner. 2012. Eyes as windows to the soul: Gazing behavior is related to personality. *Journal of Research in Personality* 46, 2 (2012), 147–156.

[60] Evan F Risko, Nicola C Anderson, Sophie Lanthier, and Alan Kingstone. 2012. Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition* 122, 1 (2012), 86–90.

[61] Rachid Rhyad Saboundji, Kinga Bettina Faragó, and Violetta Firyaridi. 2024. Prediction of Attention Groups and Big Five Personality Traits from Gaze Features Collected from an Outlier Search Game. *Journal of Imaging* 10, 10 (2024).

[62] Mark Sanderson and Susan Dumais. 2007. Examining repetition in user search behavior. In *European Conference on Information Retrieval*. Springer, 597–604.

[63] Samer Muthana Sarsam, Hosam Al-Samarraie, and Ahmed Ibrahim Alzahrani. 2023. Influence of personality traits on users' viewing behaviour. *Journal of Information Science* 49, 1 (2023), 233–247.

[64] Muhammad Shoaib, Stephan Bosch, Hans Scholten, Paul J. M. Havinga, and Ozlem Durmaz Incel. 2015. Towards detection of bad habits by fusing smartphone and smartwatch sensors. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops).* 591–596.

[65] Yoshiyuki Shoji, Kenro Aihara, Noriko Kando, Yuta Nakashima, Hiroaki Ohshima, Shio Takidaira, Masaki Ueta, Takehiro Yamamoto, and Yusuke Yamamoto. 2021. Museum experience into a souvenir: generating memorable postcards from guide device behavior log. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL).* IEEE, 120–129.

[66] Paul J Silvia and Alexander P Christensen. 2020. Looking up at the curious personality: Individual differences in curiosity and openness to experience. *Current Opinion in Behavioral Sciences* 35 (2020), 1–6.

[67] Sanjay Srivastava, Oliver P John, Samuel D Gosling, and Jeff Potter. 2003. Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of personality and social psychology* 84, 5 (2003), 1041.

[68] Ronnie Taib, Shlomo Berkovsky, Irena Koprinska, Eileen Wang, Yucheng Zeng, and Jingjie Li. 2020. Personality sensing: detection of personality traits using physiological responses to image and video stimuli. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 3 (2020), 1–32.

[69] Elina Tsigeman, Viktoria Zemliak, Maxim Likhanov, Kostas A Papageorgiou, and Yulia Kovas. 2024. AI can see you: Machiavellianism and extraversion are reflected in eye-movements. *Plos one* 19, 8 (2024), e0308631.

[70] Zhiguang Wang, Weizhong Yan, and Tim Oates. 2017. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN).* IEEE, 1578–1585.

[71] Matthew B Winn, Dorothea Wendt, Thomas Koelewijn, and Stefanie E Kuchinsky. 2018. Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want To Get Started. *Trends in hearing* 22 (2018), 1–32.

[72] Callum Woods, Zhiyuan Luo, Dawn Watling, and Szonya Durant. 2022. Twenty seconds of visual behaviour on social media gives insight into personality. *Scientific Reports* 12, 1 (2022), 1178.

[73] David W-L Wu, Walter F Bischof, Nicola C Anderson, Tanya Jakobsen, and Alan Kingstone. 2014. The influence of personality on social attention. *Personality and Individual Differences* 60 (2014), 25–29.

[74] Zahra Yusefi Hafshejani, Marjan Kaedi, and Afsaneh Fatemi. 2018. Improving sparsity and new user problems in collaborative filtering by clustering the personality factors. *Electronic Commerce Research* 18, 4 (2018), 813–836.

[75] Ramtin Zargari Marandi, Pascal Madeleine, Øyvind Omland, Nicolas Vuillerme, and Afshin Samani. 2018. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Scientific reports* 8, 1 (2018), 13148.

[76] Saber Zerhoudi, Adam Roegiest, and Johanne R. Trippas. 2026. Simulation of Interactive Information Retrieval: A Guided Tour. In *Proceedings of the ACM Conference on Information Interaction and Retrieval (CHIIR'26).*