

Learning Personalised Human Internal Cognition from External Expressive Behaviours for Real Personality Recognition

Xiangyu Kong¹, Hengde Zhu², Haoqin Sun³, Zhihao Guo⁴, Jiayan Gu², Xinyi Ni¹, Wei Zhang⁵, Shizhe Liu⁶, Siyang Song^{1*},

¹ Department of Computer Science, University of Exeter

² School of Computing and Mathematical Sciences, University of Leicester

³ College of Computer Science, Nankai University

⁴ The Manchester Metropolitan University

⁵ School of Software Technology, Zhejiang University

⁶ Department of Computer Science, University of Oxford

Abstract

Automatic real personality recognition (RPR) aims to evaluate human real personality traits from their expressive behaviours. However, most existing solutions generally act as external observers to infer observers' personality impressions based on target individuals' expressive behaviours, which significantly deviate from their real personalities and consistently lead to inferior recognition performance. Inspired by the association between real personality and human internal cognition underlying the generation of expressive behaviours, we propose a novel RPR approach that efficiently simulates personalised internal cognition from easy-accessible external short audio-visual behaviours expressed by the target individual. The simulated personalised cognition, represented as a set of network weights that enforce the personalised network to reproduce the individual-specific facial reactions, is further encoded as a novel graph containing two-dimensional node and edge feature matrices, with a novel 2D Graph Neural Network (2D-GNN) proposed for inferring real personality traits from it. To simulate real personality-related cognition, an end-to-end strategy is designed to jointly train our cognition simulation, 2D graph construction, and personality recognition modules. Experiments show our approach's effectiveness in capturing real personality traits with superior computational efficiency. Our code is provided in Supplementary Material.

Introduction

Human personality represents distinctive patterns of thoughts, emotions and behaviours defining individuals (Corr and Matthews 2020), shaped by intricate cognitive processes (VandenBos 2007). While accurately understanding human real personality is crucial for various real-world applications such as personalizing human-computer interaction (HCI) systems (Norman 2017), job recruitment (Cole et al. 2009) and mental illness diagnosis (Jaiswal, Song, and Valstar 2019), traditional self-reported questionnaires or projective tests-based personality recognition are frequently limited by subjective self-assessments and psychologists' inconsistent interpretations (Tuber 2012; McDonald 2008).

*Corresponding author.

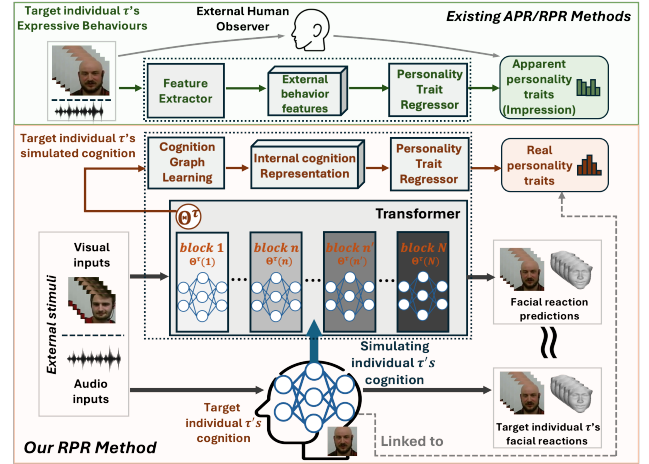


Figure 1: **Upper:** Existing methods play the role of an **external observer** to directly infer **personality impressions from the individual's external behaviours**. **Lower:** Our novel method simulates personalised internal cognition Θ^τ of individual τ by enforcing a personalised network FRG_{Θ^τ} to reproduce τ 's facial reactions in response to audio-visual inputs, and then infers **real personality from the simulated internal cognition** Θ^τ .

Since personality can be partially reflected by human expressive behaviours (e.g., facial expressions and speeches (Naumann et al. 2009)), previous studies have developed automatic, objective and repeatable models according to individuals' non-verbal facial or/and audio behaviours (Li et al. 2020; Song et al. 2023a; Curto et al. 2021; Zhang, Peng, and Winkler 2022; Dodd, Song, and Gunes 2023; Salam et al. 2022; Suman et al. 2022). These approaches can be categorized into two types (Vinciarelli and Mohammadi 2014): (1) apparent personality recognition (APR) predicting human observers' impressions on the target individual; and (2) real personality recognition (RPR) predicting the individual's real personality traits related to their internal cognition (Ajzen 2012; Daly and McCroskey 1987). Despite different goals, these APR (Ryumina et al. 2024a; Song et al. 2023a;

Li et al. 2020) and RPR (Salam et al. 2022; Palmero et al. 2021)) approaches generally infer personality directly from the target individual’s observable audio-facial behaviours, which play the role of external observers to provide *impressions that are largely different from real personality traits*, i.e., they are theoretically designed to predict apparent personality traits. This is evidenced by the poor performances of existing APR and RPR models in real personality recognition (Liao, Song, and Gunes 2024).

Inspired by the insights that human internal cognition (Farmer and Matlin 2019), which is intrinsically linked to real personality, shapes their external speech and facial behaviours (Corr and Matthews 2020; Song et al. 2023b; Shao et al. 2021), this paper hypothesises that existing deep neural networks (DNNs), designed with the inspiration from biological neural systems within human brain (Yegnaranayana 2009; Krogh 2008), can *partially approximate human cognition* by computationally task-performing based on human-like input-output mappings (inspired by the ‘Weak AI’ hypothesis (Flowers 2019)). In this sense, we innovatively propose to learn a set of personalised weights that enforces a pre-defined network to reproduce personalised facial reactions (output) expressed by the target individual when perceiving the same external stimulus (input). Consequently, the learned personalised weights would moderately simulate the individual’s internal cognition, which is further leveraged to infer real personality, i.e., our approach is theoretically designed to predict real personality (illustrated in Fig. 1). Unlike prior time-consuming Neural Architecture Search-based cognition simulation strategies (Song et al. 2023b; Shao et al. 2021) that require personalised network exploring/training on a long dyadic interaction clip pair (i.e., not recordable by the target individual alone) at the inference stage, our approach only performs efficient forward inference of the personalised weights based on an easy-accessible and short audio-facial clip of the individual without training.

To facilitate inferring real personality from human cognition, our approach further encodes the simulated personalised cognition to a processable cognition graph representation, allowing crucial structural information within network weights to be retained, which is finally fed to our proposed 2D Graph Neural Network (2D-GNN) for RPR. The main contributions and novelties (more details are in Supplementary Material) are summarised as follows:

- We propose a novel RPR framework that end-to-end encodes the target individual’s external audio-facial behaviours into a 2D graph representation representing this individual’s internal cognition, where an efficient yet effective personalised cognition simulation strategy and a 2D Graph Neural Network are proposed.
- We propose a novel end-to-end training strategy that jointly optimises the cognition simulation and real personality prediction modules, enabling our framework to effectively simulate real personality-related personalised cognition cues from easy-accessible human external audio-visual behaviours.
- Experiments not only demonstrate that our framework is the new state-of-the-art in recognising real personal-

ity traits with significant advantages over methods of traditional framework and 2,000 times inference efficiency compared to the previous cognition simulation strategy.

Related Work

Automatic personality recognition: Audio-visual automatic personality recognition methods, encompassing both apparent personality recognition (Gürpınar, Kaya, and Salah 2016; Li et al. 2020; Curto et al. 2021; Suman et al. 2022; Dodd, Song, and Gunes 2023) and real personality recognition (Salam et al. 2022; Palmero et al. 2021), have been widely developed based on the pipeline which directly extracts expressive behavioural features from human face (Zhang, Peng, and Winkler 2022; Song et al. 2023a; Ventura, Masip, and Lapedriza 2017), body (Romeo et al. 2021; Sonlu et al. 2024), speech (Liu et al. 2020; Leekha et al. 2024) or their combinations (Salam et al. 2022; Li et al. 2020; Ghassemi et al. 2023). A typical example is the CR-Net (Li et al. 2020) which leverages a ResNet-34 (He et al. 2016) to directly extract personality features from human visual, audio and text behaviours. Alternatively, a recent approach (Song et al. 2023b; Shao et al. 2021) attempts to simulate human internal cognition via their expressive behaviours (i.e., exploring a personalised network with a unique architecture to represent each individual’s cognition), and then infers real personality traits based on the simulated personalised cognition. Consequently, this approach is hindered by extremely high computational costs of personalised network exploration and dependency on long-duration dyadic interaction data at the inference stage.

Human cognition and real personality: Human cognition refers to human mental processes such as understanding and planning. Various psychological studies (Corr and Matthews 2020; Soubelet and Salthouse 2011) claimed that human cognition is strongly associated with their real personality traits. Specifically, Extraversion and Neuroticism traits are typically linked with specific biases in human cognitive functions, contributing to either extroverted social behaviours or increased threat awareness for high neurotic individuals (Boyle, Matthews, and Saklofske 2008). Meanwhile, some neuroscience studies (Adelstein et al. 2011; DeYoung et al. 2010) suggest that distinct brain cognitive patterns are associated with specific personality traits, emphasising how real personality modulates both emotional and cognitive responses to various external stimuli.

Methodology

Hypothesis: Previous studies claimed that deep neural networks (DNNs) can link certain cognitive functions (e.g., visual object recognition) to human brain (Kriegeskorte and Douglas 2018; Spoerer, McClure, and Kriegeskorte 2017), while some DNNs can approximate human cognition to perform realistic reasoning from languages (Binz et al. 2025) or multi-modal sensory inputs (Ye, Wang, and Wang 2018; Panella, Fragonara, and Tsourdos 2021; Song et al. 2023b). Given that Transformer mirrors hierarchical information processing functions corresponding to various human brain regions (Yang, Gee, and Shi 2024), we hypothesise that our

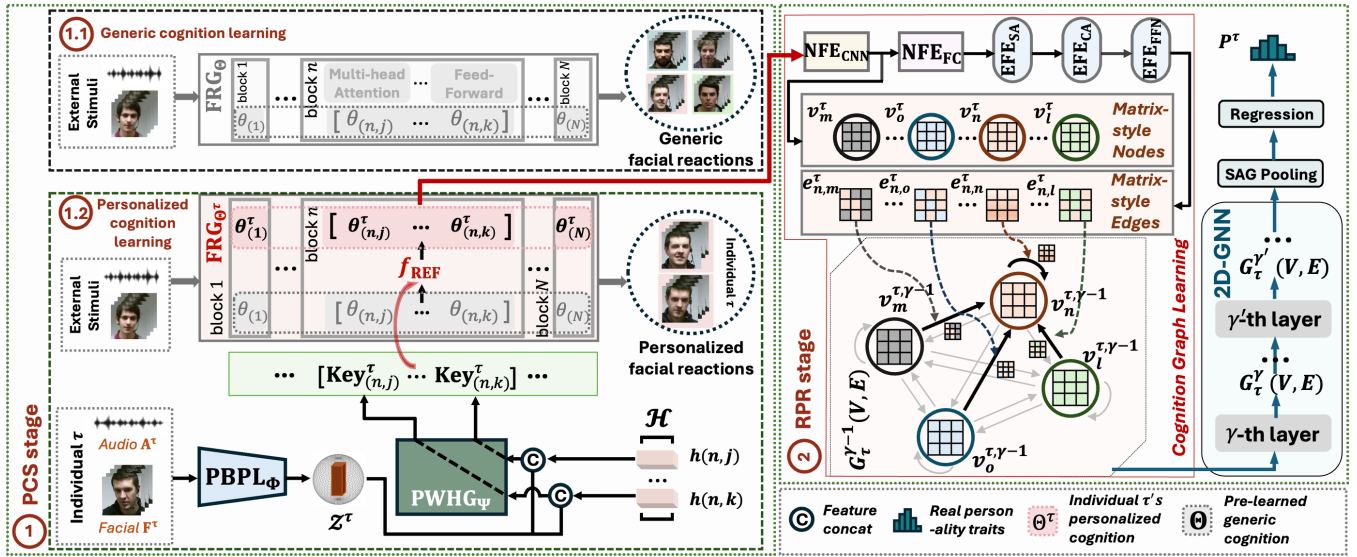


Figure 2: The pipeline of our method, which consists of two main stages: (1) Personalised Human Cognition Simulation (PCS) including (1.1) generic human cognition learning and (1.2) personalised cognition learning; and (2) novel cognition graph representation learning and 2D Graph Neural Network (2D-GNN)-based real personality recognition (detailed in **Methodology overview** section).

model composed of stacked transformer blocks is capable of approximately simulating human cognition.

Methodology overview: As depicted in Fig. 2, our approach is structured in two stages to recognise the target individual τ 's real personality traits from an external expressive behaviour expressed by τ . The **Personalised Cognition Simulation (PCS)** stage first learns a generic human cognition Θ (a set of network weights) defining a transformer-based generic facial reaction generator FRG_{Θ} capable of generating appropriate facial reactions in response to each input (perceived) stimuli, which are plausible to be expressed by humans with varying real personalities (Fig. 2 (1.1)). Then, an encoder-decoder (PBPL and PWHG in Fig. 2) is developed to generate a personalised key Key^{τ} (e.g., a set of weight offsets that have the same form as Θ) from the individual τ 's short audio and facial behaviour $\{A^{\tau}, F^{\tau}\}$, which refines Θ as Θ^{τ} (a set of network weights) representing the individual τ 's personalised cognition through the operation f_{REF} (Fig. 2 (1.2)) as:

$$\Theta^{\tau} = f_{\text{REF}}(\text{Key}^{\tau}, \Theta) \quad (1)$$

where different REF operations are evaluated in Table 3. Here, Θ^{τ} is learned to re-define the generic generator FRG_{Θ} as a personalised facial reaction generator $\text{FRG}_{\Theta^{\tau}}$, enforcing it to reproduce only τ 's personalised facial reactions. Since the pre-obtained generic cognition Θ can always be reused, the Θ^{τ} of any individual τ can be efficiently achieved from a short audio-visual behaviour expressed by τ .

Subsequently, the **Real Personality Recognition (RPR)** stage starts with a **Cognition Graph Learning (CGL)** module to innovatively encode Θ^{τ} (a set of personalised weights within the generator $\text{FRG}_{\Theta^{\tau}}$) that *cannot be directly processed by standard deep learning (DL) models* as a process-

able personalised cognition graph $G^{\tau}(V, E)$ as:

$$G^{\tau}(V, E) = \text{CGL}(\Theta^{\tau}) \quad (2)$$

where $G^{\tau}(V, E)$ is made up of a set of two-dimensional matrix-valued nodes V and edges E that retain crucial structural cues of $\text{FRG}_{\Theta^{\tau}}$'s weights Θ^{τ} as well as their relationships. Finally, a novel 2D-GNN (2D-GNN) that can process nodes and edges characterized by 2D feature matrices within $G^{\tau}(V, E)$ is proposed to predict real personality traits of τ from the cognition graph $G^{\tau}(V, E)$ (Fig. 2 (2)) as:

$$P^{\tau} = \text{2D-GNN}_{\varphi}(G^{\tau}(V, E)) \quad (3)$$

In this paper, we propose a novel joint training strategy to optimise the entire framework including $\text{FRG}_{\Theta^{\tau}}$, PBPL, PWHG, CGL, and 2D-GNN in an end-to-end manner, allowing the well-trained PBPL and PWHG to specifically simulate real personality-related personalised cognition from the target individual's external behaviours.

Personalised Cognition Simulation

Since the amount of behavioural data exhibited by a single individual may be limited while individuals of different personalities sharing certain common cognitive processes, the **PCS stage** starts with pre-learning of generic human cognition Θ of varying individuals by training FRG to reproduce facial reactions expressed by different individuals in response to input audio-facial behaviours. This allows an efficient inference stage to be achieved by only learning a set of personalised Key^{τ} from relatively small amounts of behaviours expressed by the target individual τ , e.g., a short audio-facial clip $\{A^{\tau}, F^{\tau}\}$ that can be easily recorded in non-invasive manners, where Key^{τ} is expected to refine the pre-obtained Θ as the target personalised cognition Θ^{τ} ,

Specifically, we first train the generic facial reaction generator FRG_Θ (i.e., simulating generic cognition) using dyadic audio-facial human-human interaction clips expressed by various pairs of individuals, with the objective to output appropriate facial reactions $\mathcal{R} = \{R_1, R_2, \dots, R_L\}$ expressed by individuals of different personalities in response to each input human behaviour $\{A^s, F^s\}$ as:

$$\mathcal{R} = \text{FRG}_\Theta(A^s, F^s) \quad (4)$$

where Θ is learned as a set of network weights. To validate our hypothesis and ensure reproducibility, the FRG_Θ simply stacks N transformer blocks with each consisting of a self-attention, a cross-attention, and a feed-forward layer containing two FC layers, which includes K weights (i.e., 2D matrices for all linear layers).

Then, the **Personalised Behaviour Pattern Learning (PBPL)** module learns a latent representation $\mathcal{Z}^\tau$ to describe the individual τ 's *personalised external behaviour patterns* $\mathcal{Z}^\tau$ (a feature vector) from individual's *short audio-facial behaviour clip* $\{A^\tau, F^\tau\}$ as:

$$\mathcal{Z}^\tau = \text{PBPL}(A^\tau, F^\tau) \quad (5)$$

where PBPL is a transformer-based encoder. Subsequently, a **Personalised Weight HyperGenerator (PWHG)** consisting of FC layers further projects $\mathcal{Z}^\tau$ to the personalised key Key^τ (e.g., a set of weight offsets) to refine the generic cognition Θ as the personalised cognition Θ^τ as Eq. 1. To efficiently learn $K \times N$ matrices in Θ (K matrices for parameterizing all linear layers in each of the N transformer blocks), our PWHG module samples $N \times K$ learnable chunk embeddings $\mathcal{H} = \{h_{n,k} | n = 1, \dots, N; k = 1, \dots, K\}$ from the standard Gaussian distribution, and then concatenates each $h_{n,k}$ with the learned $\mathcal{Z}^\tau$. This results in $K \times N$ different concatenated embeddings, which are then projected to $K \times N$ matrices as Key^τ via a shared PWHG module as:

$$\begin{aligned} \text{Key}^\tau &= \{[\text{PWHG}(\mathcal{Z}^\tau, h_{n,k})]\}_{n=1:N, k=1:K} \\ &= \{[\text{Key}_{n,k}^\tau]\}_{n=1:N, k=1:K} \end{aligned} \quad (6)$$

This way, the PWHG only includes one shared projector for generating all weight matrices, resulting in approximately $K \times N$ -fold reduction in the number of learnable weights compared to directly using $K \times N$ independent projectors.

At the inference stage, since the pre-learned Θ can always be re-used without training (i.e., the competitor (Song et al. 2023b) requires the time-consuming personalised network exploration for every individual during inference), our approach is efficient as it only involves forwardly propagating the given audio-facial clip A^τ, F^τ to refine Θ as Θ^τ .

Cognition-based real personality recognition

Since the simulated personalised cognition Θ^τ is a set of unorganised weights that are difficult to be directly processed by standard DL models for RPR, the RPR stage encodes the Θ^τ as an organised and processable graph G^τ containing node and edge feature matrices. Then, a novel 2D Graph Neural Network (2D-GNN) is proposed to infer real personality traits from G^τ .

Cognition graph representation (nodes and edges) learning: Given Θ^τ , the RPR stage applies a **Cognition Graph Learning (CGL)** module consisting of a node feature encoding (NFE) block and an edge feature encoding (EFE) block to encode the weight matrices parameterizing all linear layers in N transformer blocks of FRG_{Θ^τ} to a personalised cognition graph $G^\tau(V, E)$, where $V = \{v_1, \dots, v_N\}$ denotes N nodes and $E = \{e_{n,m} | n, m = 1, \dots, N\}$ denotes a set of directed edges connecting all node pairs. Specifically, the **NFE** block encodes the K equal-size weights $\Theta_n^\tau = [\Theta_{n,1}^\tau, \dots, \Theta_{n,K}^\tau]$ of the n -th transformer block into a single node feature $v_n^\tau \in \mathbb{R}^{d_v \times d_v}$ as:

$$v_n^\tau = \text{NFE}_{\text{FC}}(\text{NFE}_{\text{CNN}}(\Theta_n^\tau)) \quad (7)$$

It first processes every $\Theta_n^\tau \in \mathbb{R}^{K \times d \times d}$ through a CNN extractor NFE_{CNN} to obtain the feature representation η_n^τ , which is then processed by a FC layer NFE_{FC} to yield the matrix-valued node v_n^τ (i.e., weighted sum of K layer-level weights). This way, the final obtained G^τ preserves crucial structural information within weights, avoiding information loss/distortion when reducing them to vectors (evaluated in Table 5). To comprehensively explore real personality-related relationships among the simulated cognitive processes (i.e., transformer blocks' weights represented by graph node representations), our **EFE** block learns a pair of edge feature matrices $e_{n,m}^\tau$ and $e_{m,n}^\tau$ between every pair of nodes v_n^τ and v_m^τ . This strategy is proposed not only because it can explore all potential personality-related relationship cues among transformer blocks (i.e., the exact relationship among nodes remains unclear), but also the widely-used single-value edge features fail to comprehensively model complex relationships between a pair of node features unless they are linearly dependent. Specifically, our EFE applies a self-attention (EFE_{SA}) and a cross-attention (EFE_{CA}) to emphasise real personality-related relationships shared by every pair of node features, followed by a feed-forward network (EFE_{FFN}) for further refinement. These result in a pair of edge features $e_{n,m}^\tau \in \mathbb{R}^{d_v \times d_v}$ and $e_{m,n}^\tau \in \mathbb{R}^{d_v \times d_v}$ as:

$$\begin{aligned} e_{n,m}^\tau &= \text{EFE}(\eta_n^\tau, \eta_m^\tau) \\ &= \text{EFE}_{\text{FFN}}(\text{EFE}_{\text{CA}}(\text{EFE}_{\text{SA}}(\eta_n^\tau), \text{EFE}_{\text{SA}}(\eta_m^\tau))) \end{aligned} \quad (8)$$

More model and implementation details of the proposed NFE and EFE are provided in the Supplementary Material.

2D Graph Neural Network (2D-GNN): While existing standard GNNs are limited to processing graphs with vector-based node and edge features, our novel 2D-GNN directly processes node feature matrices V and edge feature matrices E within the learned cognition graph $G^\tau(V, E)$ to infer τ 's real personality traits $\mathcal{P}_{\text{traits}}^\tau$. Specifically, the γ -th 2D-GNN layer updates its input node feature $v_n^{\tau, \gamma-1} \in \mathbb{R}^{d_v \times d_v}$ within the graph $G^\tau(V, E)^{\gamma-1}$ as an node feature $v_n^{\tau, \gamma} \in \mathbb{R}^{d_v \times d_v}$ of the same size as:

$$\begin{aligned} v_n^{\tau, \gamma} &= \text{ReLU}\left(\frac{1}{d_n} \sum_{m \in \mathcal{N}(n)} A_{n,m}^\tau \cdot \hat{e}_{n,m}^\tau \odot \hat{v}_n^{\tau, \gamma-1}\right), \\ \hat{e}_{n,m}^\tau &= f_e^{(\gamma-1)}(e_{n,m}^\tau), \quad \hat{v}_n^{\tau, \gamma-1} = f_v^{(\gamma-1)}(v_n^{\tau, \gamma-1}) \end{aligned} \quad (9)$$

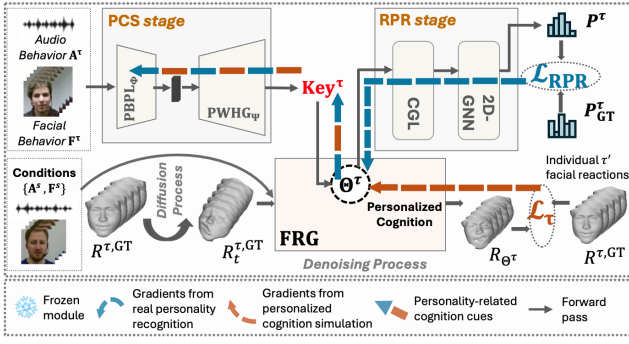


Figure 3: Our end-to-end joint training strategy is driven by two loss functions: (1) $\mathcal{L}_\tau$ compares the generated facial reaction R_{Θ^τ} with the GT personalised real facial reaction $R^{\tau,GT}$. The gradients are backpropagated to optimise PWHG_Ψ and PBPL_Φ ; (2) $\mathcal{L}_{\text{RPR}}$ compares the personality prediction $\mathcal{P}^\tau$ with the GT real personality traits $\mathcal{P}_{\text{GT}}^\tau$. The gradients are back-propagated to train 2D-GNN, CGL, PWHG_Ψ and PBPL_Φ . This way, both losses jointly enforce our PCS to simulate personality-related cognition from input audio-visual behaviours.

where $f_e^{(\gamma-1)}$ and $f_v^{(\gamma-1)}$ are linear projections; d_n is the degree of node $v_n^{\tau,\gamma-1}$ for normalization; $\odot$ represents Hadamard product, $A_{n,m}^\tau \in \{0,1\}$ is binary adjacency indicator and $\mathcal{N}(n)$ denotes the $v_n^{\tau,\gamma-1}$'s neighboring nodes. This way, each node $v_n^{\tau,\gamma}$ is obtained by considering both its previous states $v_n^{\tau,\gamma-1}$ and the messages passed from its neighbouring nodes $\mathcal{N}(n)$ via the corresponding edge feature matrices. Finally, a self-attention graph pooling (Lee, Lee, and Kang 2019) integrates all updated node features, based on which three FC layers equipped with ReLU and dropout are finally employed to simultaneously infer five real personality traits.

End-to-end joint training strategy

We pre-train the pre-defined architecture FRG as a generic facial reaction generator FRG_Θ in a diffusion manner with the objective loss function defined as:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t \in [1,T], R_t^{\text{GT}} \sim q_t} \left[\|R_t^{\text{GT}} - R_{\Theta^\tau}\|_2^2 \right] \quad (10)$$

$$R_{\Theta^\tau} = \text{FRG}_\Theta(R_t^{\text{GT}}, t, \{A^s, F^s\})$$

where R_t^{GT} is obtained by adding random Gaussian noises to the ground-truth (GT) facial reaction R^{GT} via a forward diffusion process q_t (Ho, Jain, and Abbeel 2020); t is the diffusion step controlling the noise level; and $\{A^s, F^s\}$ are the input audio and facial behaviours) that triggered the corresponding real facial reaction R^{GT} . Since R^{GT} is expressed by individuals with different personalities, FRG_Θ learns to simulate generic human cognition for generating various appropriate facial reactions that can be expressed by individuals of different personalities.

Based on the pre-trained FRG_Θ , our end-to-end joint training strategy is achieved by two loss functions. As illustrated in Fig. 3, the **first loss** $\mathcal{L}_\tau$ compares the facial reaction

R_{Θ^τ} generated by the refined personalised facial reaction generator FRG_{Θ^τ} with the corresponding real personalised appropriate facial reaction $R^{\tau,GT}$ expressed by τ (i.e., any facial reaction expressed by τ that is appropriate in response to the input A^s, F^s) (Zhu et al. 2024) as:

$$\mathcal{L}_\tau = \mathbb{E}_{t \in [1,T], R_t^{\tau,GT} \sim q_t} \left[\|R_t^{\tau,GT} - R_{\Theta^\tau}\|_2^2 \right] \quad (11)$$

$$R_{\Theta^\tau} = \text{FRG}_{\Theta^\tau}(R_t^{\tau,GT}, t, \{A^s, F^s\})$$

where $R_t^{\tau,GT}$ is obtained by adding random Gaussian noises to $R^{\tau,GT}$ via a forward diffusion process q_t . During the training, the FRG_Θ are kept frozen, while the PBPL_Φ and PWHG_Ψ modules generate the personalised key Key^τ to refine (i.e., via a differentiable adding operation) the weights of FRG_Θ as FRG_{Θ^τ} that aims at recovering personalised facial reaction $R^{\tau,GT}$ from $R_t^{\tau,GT}$. This way, the gradients computed from the loss between the generated personalised facial reaction and the ground-truth $R^{\tau,GT}$ are back-propagated through the frozen FRG_Θ and Key_τ to optimise PBPL_Φ and PWHG_Ψ . Here, the gradient of $\mathcal{L}_\tau$ with respect to Key_τ is:

$$\xi_{n,k}^\tau = \sigma(\Theta_{n,k}^\tau \xi_{n,k-1}^\tau + b_{n,k}),$$

$$\frac{\partial \mathcal{L}_\tau}{\partial \text{Key}_{n,k}^\tau} = \frac{\partial \mathcal{L}_\tau}{\partial \xi_{n,k}^\tau} \cdot \frac{\partial \xi_{n,k}^\tau}{\partial \Theta_{n,k}^\tau} \cdot \frac{\partial \Theta_{n,k}^\tau}{\partial \text{Key}_{n,k}^\tau},$$

$$= \frac{\partial \mathcal{L}_\tau}{\partial \xi_{n,k}^\tau} \cdot \sigma' \cdot \xi_{n,k-1}^\tau \cdot f'_{\text{REF}}(\Theta, \text{Key}_{n,k}^\tau) \quad (12)$$

where we denote $\xi_{n,k}^\tau$ as the intermediate activation generated from the k -th layer of the n -th block during the forward propagation of FRG_{Θ^τ} (i.e., comprising the weight parameter $\Theta_{n,k}^\tau$ and the bias parameter $b_{n,k}$) while σ' denoting the derivative of the activation function σ . Since FRG_{Θ^τ} is produced via the refinement function (defined in Eq. 1) based on the personalised $\text{Key}_{n,k}^\tau$, $f'_{\text{REF}}(\Theta, \text{Key}_{n,k}^\tau)$ represents its derivative with respect to $\text{Key}_{n,k}^\tau$.

Meanwhile, the **second loss** $\mathcal{L}_{\text{RPR}}$ compares the personality traits $\mathcal{P}^\tau$ predicted by our 2D-GNN and the ground-truth real personality traits $\mathcal{P}_{\text{GT}}^\tau$ as:

$$\mathcal{L}_{\text{RPR}} = \|\mathcal{P}^\tau - \mathcal{P}_{\text{GT}}^\tau\|^2 \quad (13)$$

More training and loss function details are provided in Supplementary Material.

Experiments

Experimental settings

Datasets: We evaluate our method on two publicly available audio-visual datasets rigorously labeled with clip-level self-reported Big-Five real personality traits (Extraversion (Ex), Agreeableness (Ag), Openness (Op), Conscientiousness (Co), and Neuroticism (Ne)), including: (1) NoXI (Cafaro et al. 2017) is a multilingual dyadic interaction dataset designed to study adaptive behaviours in ‘‘Expert-Novice’’ roles through video conferencing, encompassing 84 sessions across various topics; and (2) UDIVA (Palmero et al. 2021), comprising 188 in-person dyadic interactions from 147 participants recorded under four different tasks.

UDIVA - MSE ($\downarrow$)							UDIVA - CCC ($\uparrow$)						
Methods	Open	Cons	Extrav	Agree	Neuro	Avg.	Methods	Open	Cons	Extrav	Agree	Neuro	Avg.
TPN (Yang et al. 2020)	0.9259	0.7596	1.2524	0.9458	1.4147	1.0597	TPN (Yang et al. 2020)	0.0448	0.0438	0.0287	-0.0177	-0.0281	0.0125
DCC (Güçlütürk et al. 2016)	0.9637	0.6709	1.5770	1.0364	1.5126	1.0801	DCC (Güçlütürk et al. 2016)	0.0648	0.2424	0.0349	0.0305	0.0009	0.0743
CAM-DAN+ (Ventura, Masip, and Lapedriza 2017)	1.0223	0.5471	1.4800	0.8446	1.1493	1.0087	CAM-DAN+ (Ventura, Masip, and Lapedriza 2017)	0.0336	0.3359	0.0270	0.2014	0.2709	0.1738
DAN (Wei et al. 2018)	1.0841	1.1174	2.0979	1.4846	1.2400	1.4048	DAN (Wei et al. 2018)	0.0009	0.0087	0.0061	0.0035	-0.0016	0.0035
DAN-MFCC (Wei et al. 2018)	1.1657	0.7358	2.0523	1.1006	1.0745	1.2258	DAN-MFCC (Wei et al. 2018)	0.0008	0.1154	0.0132	-0.0022	0.1219	0.0498
PersEmoN (Zhang, Peng, and Winkler 2022)	0.9034	0.6985	1.4097	0.9446	1.1469	1.0206	PersEmoN (Zhang, Peng, and Winkler 2022)	-0.0095	0.0133	0.0045	0.0058	0.0045	0.0037
CR-Net (Li et al. 2020)	1.0985	0.6877	1.3079	0.9099	1.4140	1.0834	CR-Net (Li et al. 2020)	0.0998	0.1788	0.1158	0.2168	0.0449	0.1312
Amb-Fac (Suman et al. 2022)	0.8529	0.6394	1.4212	0.9334	1.1410	0.9976	Amb-Fac (Suman et al. 2022)	0.0532	0.1941	0.0442	0.0453	0.0204	0.0714
OCEAN-AI (Ryumina et al. 2024b)	0.9391	0.6675	1.4956	0.9588	1.1590	1.0440	OCEAN-AI (Ryumina et al. 2024b)	-0.0348	0.0468	0.0302	0.0397	0.0041	0.0171
Ours	0.7492	1.1114	0.7166	0.7829	0.7026	0.8125	Ours	0.2257	-0.0345	0.2208	0.2988	0.4431	0.2308
NoXI - MSE ($\downarrow$)							NoXI - PCC ($\uparrow$)						
Methods	Open	Cons	Extrav	Agree	Neuro	Avg.	Methods	Open	Cons	Extrav	Agree	Neuro	Avg.
DCC (Güçlütürk et al. 2016)	0.112	0.755	0.640	0.082	0.102	0.338	Spectral (Song et al. 2022)	0.135	0.246	0.265	0.192	0.277	0.223
NJU-LAMDA (Wei et al. 2018)	0.099	0.135	0.069	0.075	0.210	0.118	CR-Net (Li et al. 2020)	0.181	0.271	0.301	0.177	0.325	0.251
VAT (Girdhar et al. 2019)	0.186	0.081	0.088	0.068	0.152	0.115	VAT (Girdhar et al. 2019)	-0.177	0.484	0.131	0.448	0.187	0.215
Amb-Fac (Suman et al. 2022)	0.094	0.047	0.035	0.038	0.049	0.053	Amb-Fac (Suman et al. 2022)	-0.127	0.109	0.059	0.107	0.050	0.040
OCEAN-AI (Ryumina et al. 2024b)	0.081	0.046	0.040	0.043	0.044	0.051	OCEAN-AI (Ryumina et al. 2024b)	-0.039	0.305	0.281	0.331	0.201	0.216
Dyadformer (Curto et al. 2021)	0.063	0.049	0.044	0.048	0.050	0.051	Dyadformer (Curto et al. 2021)	0.128	0.317	0.049	0.373	0.253	0.224
P-NAS (Song et al. 2023b)	-	-	-	-	-	-	P-NAS (Song et al. 2023b)	0.189	0.376	0.420	0.289	0.481	0.351
Ours	0.074	0.052	0.024	0.037	0.053	0.048	Ours	0.099	0.520	0.424	0.515	0.061	0.324

Table 1: Comparison between our approach with most recent audio-visual personality recognition approaches on UDIVA and NoXI datasets, where UDIVA experiments are conducted following the widely-used benchmark (Liao, Song, and Gunes 2024), and thus (Curto et al. 2021) is not compared as it specifically considered verbal/topic contexts.

Implementation details: We apply the pre-trained wav2vec 2.0 (Baevski et al. 2020) to extract frame-level (25 fps) audio behaviour features, and a 3D Morphable Face Model (3DMM) (Blaiz and Vetter 2023; Wang et al. 2022) to extract frame-level facial coefficients. The number of transformer layers within our FRG is empirically set to $N = 9$. The default length of the audio-facial behaviour clips fed to PBPL is 10 seconds. More dataset and implementation details are provided in Supplementary Material.

Evaluation Metrics: Following (Liao, Song, and Gunes 2024), we employ Mean Squared Error (MSE), Pearson Correlation Coefficient (PCC) and Concordance Correlation Coefficient (CCC) to evaluate personality recognition performances. As previous studies did not report PCC on UDIVA or CCC on NoXI, we present our results for these metrics in Supplementary Material.

Comparison with existing approaches

Table 1 shows that our cognition simulation approach clearly outperformed existing standard personality recognition methods on UDIVA datasets, achieved 18.6% and 32.8% average MSE and CCC improvements over the previous state-of-the-arts (SOTA) (Suman et al. 2022; Ventura, Masip, and Lapedriza 2017). Our approach is particularly effective in inferring Extraversion and Agreeableness traits-related personalised cognition, as it achieved the best CC-C/MSE results for them on both datasets. We explain this as these two traits may be more related to human external expressive behaviours (Lepri et al. 2012; Graziano and Tobin 2009). For the NoXI dataset, our approach achieved the best MSE results on 3/5 OCEAN traits (i.e., Conscientiousness, Extraversion and Agreeableness) as well as the averaged MSE (5.9% relative improvements over the previous SOTA). It also achieved the second best CCC performance (29.1% relative average CCC improvement over the third best method) with small disadvantages compared to the impractical method (Song et al. 2023b) but *over 2,000 times faster inference speed* (Table 2). Importantly, our approach not only achieved clear superior personality recogni-

tion results but also experienced similar computational costs compared to existing standard personality recognition approaches. Fig. 4 shows that individuals with similar personality trait distributions (grouped by three per set) exhibit similar weight patterns learned from their audio-facial behaviours. Additionally, these individuals display comparable facial reactions to the same external stimulus (e.g., the left group shows reserved facial reactions, while the right group is more animated), highlighting the ability of our approach to capture personality-related cognition.

Methods	Inference Time (per τ)	NoXI-PCC (Avg.)
OCEAN-AI (Ryumina et al. 2024b)	1.3 minutes	0.216
DCC (Güçlütürk et al. 2016)	1.8 minutes	0.008
Dyadformer (Curto et al. 2021)	4.46 minutes	0.224
Amb-Fac (Suman et al. 2022)	3.20 minutes	0.040
P-NAS (Song et al. 2023b)	8.3 days	0.351
Ours	5.29 minutes	0.324

Table 2: Inference time of RPR models on a 10-minute audio-visual clip, where P-NAS (Song et al. 2023b) and ours are cognition-based methods, while others play the role of external observers.

Ablation studies

This section conducts a series of ablation studies on NoXI dataset. Additionally experiments are provided in Supplementary Material, comparing: (1) input modality contributions; (2) various training strategies, and investigate: (3) parameter sensitivity for key model components; (4) long-term stability of learned personality; (5) facial reaction generation results; (6) statistical significance; (7) model computational costs; and (8) model’s bias and fairness.

Influences of model components: Table 5 shows that directly inferring real personality from external behaviours in the short audio-facial clip (V1) leads to very poor performance. In contrast, with the help of FRG to simulate personalised cognition, the performance largely improved (V2), where pre-training FRG plays a crucial role (V8 vs V2 / V9 vs V3), suggesting that the simulated cognition provides

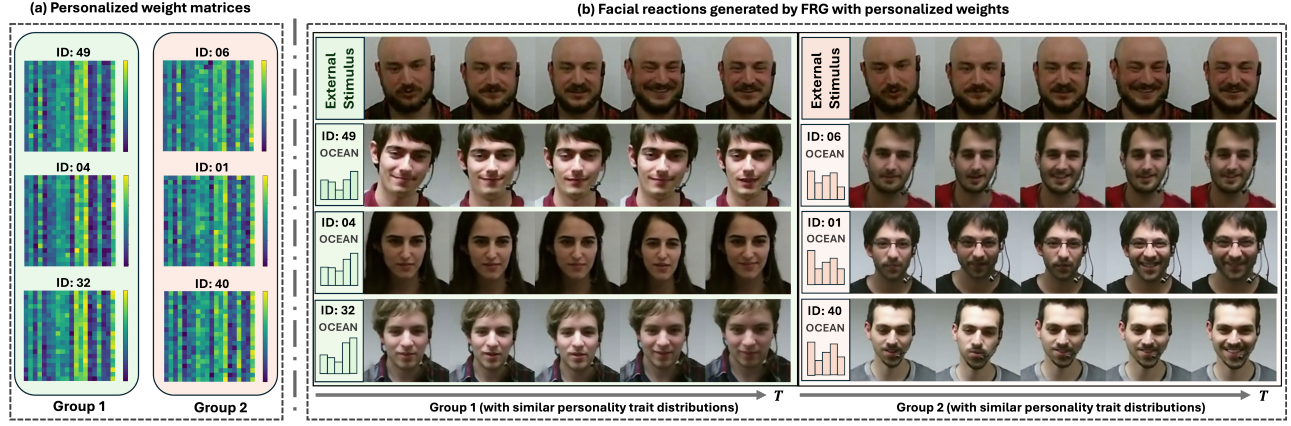


Figure 4: (a) Example personalised weight matrices learned for different individuals; (b) Individuals’ facial reactions generated by FRG with different personalised weights in response to the same stimulus.

Methods	Metric	Open	Cons	Extrav	Agree	Neuro	Avg.
Key^τ (w/o Θ)	MSE	0.075	0.067	0.034	0.064	0.046	0.057
	PCC	0.005	0.240	0.258	0.241	0.246	0.198
$\text{Key}^\tau \times \Theta$	MSE	0.058	0.055	0.034	0.037	0.053	0.048
	PCC	0.248	0.447	0.170	0.349	-0.100	0.223
$\text{Key}^\tau + \Theta$	MSE	0.074	0.052	0.024	0.037	0.053	0.048
	PCC	0.099	0.520	0.424	0.515	0.061	0.324

Table 3: Results achieved for different personalised cognition refinement strategies, where Key^τ (w/o Θ) denotes directly generating Key^τ as the personalised cognition.

Setting	Metric	Open	Cons	Extrav	Agree	Neuro	Avg.
Layer-level node (fully connected)	MSE	0.056	0.067	0.046	0.062	0.047	0.056
	PCC	0.214	0.590	0.161	0.400	0.004	0.274
Layer-level node (head-to-tail)	MSE	0.070	0.055	0.036	0.052	0.054	0.053
	PCC	0.105	0.456	0.223	0.457	0.213	0.291
Block-level node (KNN)	MSE	0.059	0.061	0.036	0.049	0.052	0.052
	PCC	0.320	0.586	0.138	0.422	0.118	0.317
Block-level node (fully connected)	MSE	0.074	0.052	0.024	0.037	0.053	0.048
	PCC	0.099	0.520	0.424	0.515	0.061	0.324

Table 4: Results achieved for different node definition/edge connection strategies. **Layer-level node**: treating each weight matrix of linear layer in each transformer block as a node; **head-to-tail**: all nodes within each transformer block are fully connected while the node of the last weight matrix within a block connects to the node of the first weight matrix in the following block; and **KNN**: empirically connecting each node to its six nearest neighbours (measured by L2 distance).

informative personality cues. Meanwhile, our transformer block-level node encoding strategy (V7 vs V5), cognition graph representation learning (node feature matrices (V5 vs V4) and edge feature matrices (V9 vs V7)), and the corresponding 2D-GNN additionally enhanced the performance. This validates that such a cognition-based graph managed to preserve crucial structural cues regarding personality-related cognition, where edge features facilitates better task-specific

	PCS stage		RPR stage			Training	Metric	Avg.
	FRG $_{\Theta}$	FRG $_{\Theta}$ Pre-training	Block-level nodes	Matrix nodes	Matrix edges	End-to-end training		
V1			✓	✓	✓		MSE PCC	0.067 0.110
V2	✓		✓	✓	✓		MSE PCC	0.060 0.179
V3	✓		✓	✓	✓	✓	MSE PCC	0.057 0.198
V4	✓	✓				✓	MSE PCC	0.054 0.158
V5	✓	✓		✓		✓	MSE PCC	0.053 0.254
V6	✓	✓	✓	✓			MSE PCC	0.052 0.197
V7	✓	✓	✓	✓		✓	MSE PCC	0.051 0.262
V8	✓	✓	✓	✓	✓		MSE PCC	0.051 0.319
V9	✓	✓	✓	✓	✓	✓	MSE PCC	0.048 0.324

Table 5: Contributions of model components. **V1**: denotes no cognition simulation involved in prediction; **V4**: applying standard vector-based graph with GCN for processing simulated cognition; **V4/V5**: encoding every weight matrix in every block to a node feature vector/matrix.

message exchanging among node features. Finally, our end-to-end training strategy consistently enhanced the performance, indicating its effectiveness in simulating personality-related cognition.

Results of different ‘Refine’ operations: Table 3 compares different strategies to project the learned Key^τ as the personalised cognition Θ^τ . It is clear that directly projecting Key^τ as Θ^τ without considering the pre-learned Θ resulted in poor performance, i.e., *it is difficult to directly obtain reliable personalised cognition from a short clip* (also evidenced by poor facial reaction generation results in Supplementary Material). Besides, simply adding Key^τ to Θ achieved superior results over applying it to weight Θ .

Results of different cognition graph node/topology settings: As compared in Table 4, when representing each linear layer’s weight matrix within each transformer block as a

node feature, the performances are not as good as our default block-level node feature setting, regardless of their topology. More importantly, our fully connected topology outperformed the topology defined by the widely-used KNN strategy, suggesting that real personality-related cognition patterns may be reflected by the combination of all simulated cognitive processes (transformer blocks/nodes).

Conclusion

This paper presents an end-to-end audio-facial personalised human cognition simulation strategy for RPR. Results show that our strategy can effectively and efficiently simulate personality-related personalised cognition while our novel cognition graph learning and 2D-GNN also positively contributed to RPR. Key limitations include the incomplete integration of nuanced verbal features and the simple network architecture, which will be addressed in our future work. Our novel framework can be potentially extended for understanding other human internal status (e.g., expressive behaviour-based mental health/emotion analysis).

References

- Adelstein, J. S.; Shehzad, Z.; Mennes, M.; DeYoung, C. G.; Zuo, X.-N.; Kelly, C.; Margulies, D. S.; Bloomfield, A.; Gray, J. R.; Castellanos, F. X.; et al. 2011. Personality is reflected in the brain's intrinsic functional architecture. *PLoS one*, 6(11): e27633.
- Ajzen, I. 2012. Attitudes and persuasion. *The Oxford handbook of personality and social psychology*, 367–393.
- Baevski, A.; Zhou, Y.; Mohamed, A.; and Auli, M. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33: 12449–12460.
- Binz, M.; Akata, E.; Bethge, M.; Brändle, F.; Callaway, F.; Coda-Forno, J.; Dayan, P.; Demircan, C.; Eckstein, M. K.; Éltető, N.; et al. 2025. A foundation model to predict and capture human cognition. *Nature*, 1–8.
- Blanz, V.; and Vetter, T. 2023. A morphable model for the synthesis of 3D faces. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 157–164.
- Boyle, G. J.; Matthews, G.; and Saklofske, D. H. 2008. Personality measurement and testing: An overview. *The SAGE handbook of personality theory and assessment*, 2: 1–26.
- Cafaro, A.; Wagner, J.; Baur, T.; Dermouche, S.; Torres, M. T.; Pelachaud, C.; André, E.; and Valstar, M. 2017. The NoXi Database: Multimodal Recordings of Mediated Novice-Expert Interactions. In *Proceedings of 19th ACM International Conference on Multimodal Interaction*, 350–359.
- Cole, M. S.; Feild, H. S.; Giles, W. F.; and Harris, S. G. 2009. Recruiters' inferences of applicant personality based on resume screening: do paper people have a personality? *Journal of Business and Psychology*, 24: 5–18.
- Corr, P. J.; and Matthews, G. 2020. *The Cambridge handbook of personality psychology*. Cambridge University Press.
- Curto, D.; Clapés, A.; Selva, J.; Smeureanu, S.; Junior, J. C. S. J.; Gallardo-Pujol, D.; Guilera, G.; Leiva, D.; Moeslund, T. B.; Escalera, S.; and Palmero, C. 2021. Dyadformer: A Multi-Modal Transformer for Long-Range Modeling of Dyadic Interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2177–2188.
- Daly, J. A.; and McCroskey, J. C. 1987. *Personality and interpersonal communication*. Sage Beverly Hills.
- DeYoung, C. G.; Hirsh, J. B.; Shane, M. S.; Papademetris, X.; Rajeevan, N.; and Gray, J. R. 2010. Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological science*, 21(6): 820–828.
- Dodd, E.; Song, S.; and Gunes, H. 2023. A framework for automatic personality recognition in dyadic interactions. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 1–8. IEEE.
- Farmer, T. A.; and Matlin, M. W. 2019. *Cognition*. John Wiley & Sons.
- Flowers, J. C. 2019. Strong and Weak AI: Deweyan Considerations. In *AAAI spring symposium: Towards conscious AI systems*, volume 2287.
- Ghassemi, S.; Zhang, T.; Van Breda, W.; Koutsoumpis, A.; Oostrom, J. K.; Holtrop, D.; and de Vries, R. E. 2023. Unsupervised multimodal learning for dependency-free personality recognition. *IEEE transactions on affective computing*.
- Girdhar, R.; Carreira, J.; Doersch, C.; and Zisserman, A. 2019. Video Action Transformer Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Graziano, W. G.; and Tobin, R. M. 2009. Agreeableness.
- Güçlütürk, Y.; Güçlü, U.; van Gerven, M. A.; and van Lier, R. 2016. Deep impression: Audiovisual deep residual networks for multimodal apparent personality trait recognition. In *Computer Vision—ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14*, 349–358. Springer.
- Gürpınar, F.; Kaya, H.; and Salah, A. A. 2016. Combining deep facial and ambient features for first impression estimation. In *Computer Vision—ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14*, 372–385. Springer.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.
- Jaiswal, S.; Song, S.; and Valstar, M. 2019. Automatic prediction of depression and anxiety from behaviour and personality attributes. In *2019 8th international conference on affective computing and intelligent interaction (acii)*, 1–7. IEEE.

- Kriegeskorte, N.; and Douglas, P. K. 2018. Cognitive computational neuroscience. *Nature neuroscience*, 21(9): 1148–1160.
- Krogh, A. 2008. What are artificial neural networks? *Nature biotechnology*, 26(2): 195–197.
- Lee, J.; Lee, I.; and Kang, J. 2019. Self-Attention Graph Pooling. In Chaudhuri, K.; and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 3734–3743. PMLR.
- Leekha, M.; Khan, S. N.; Srinivas, H.; Shah, R. R.; and Shukla, J. 2024. VyaktitvaNirdharan: Multimodal Assessment of Personality and Trait Emotional Intelligence. *IEEE Transactions on Affective Computing*.
- Lepri, B.; Subramanian, R.; Kalimeri, K.; Staiano, J.; Pineschi, F.; and Sebe, N. 2012. Connecting meeting behavior with extraversion—A systematic study. *IEEE Transactions on Affective Computing*, 3(4): 443–455.
- Li, Y.; Wan, J.; Miao, Q.; Escalera, S.; Fang, H.; Chen, H.; Qi, X.; and Guo, G. 2020. Cr-net: A deep classification-regression network for multimodal apparent personality analysis. *International Journal of Computer Vision*, 128: 2763–2780.
- Liao, R.; Song, S.; and Gunes, H. 2024. An Open-Source Benchmark of Deep Learning Models for Audio-Visual Apparent and Self-Reported Personality Recognition. *IEEE Transactions on Affective Computing*, 15(3): 1590–1607.
- Liu, Z.-T.; Rehman, A.; Wu, M.; Cao, W.-H.; and Hao, M. 2020. Speech personality recognition based on annotation classification using log-likelihood distance and extraction of essential audio features. *IEEE transactions on multimedia*, 23: 3414–3426.
- McDonald, J. D. 2008. Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1(1): 1–19.
- Naumann, L. P.; Vazire, S.; Rentfrow, P. J.; and Gosling, S. D. 2009. Personality judgments based on physical appearance. *Personality and social psychology bulletin*, 35(12): 1661–1671.
- Norman, K. L. 2017. *Cyberpsychology: An introduction to human-computer interaction*. Cambridge university press.
- Palmero, C.; Selva, J.; Smeureanu, S.; Junior, J.; Jacques, C.; Clapés, A.; Moseguí, A.; Zhang, Z.; Gallardo, D.; Guiler, G.; et al. 2021. Context-aware personality inference in dyadic scenarios: Introducing the udiva dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1–12.
- Panella, I.; Fragonara, L. Z.; and Tsourdos, A. 2021. A deep learning cognitive architecture: Towards a unified theory of cognition. In *Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 1*, 566–582. Springer.
- Romeo, M.; Hernández García, D.; Han, T.; Cangelosi, A.; and Jokinen, K. 2021. Predicting apparent personality from body language: benchmarking deep learning architectures for adaptive social human–robot interaction. *Advanced Robotics*, 35(19): 1167–1179.
- Ryumina, E.; Markitantov, M.; Ryumin, D.; and Karpov, A. 2024a. Gated Siamese Fusion Network based on multimodal deep and hand-crafted features for personality traits assessment. *Pattern Recognition Letters*, 185: 45–51.
- Ryumina, E.; Markitantov, M.; Ryumin, D.; and Karpov, A. 2024b. OCEAN-AI framework with EmoFormer cross-hemiface attention approach for personality traits assessment. *Expert Systems with Applications*, 239: 122441.
- Salam, H.; Manoranjan, V.; Jiang, J.; and Celiktutan, O. 2022. Learning personalised models for automatic self-reported personality recognition. In *Understanding Social Behavior in Dyadic and Small Group Interactions*, 53–73. PMLR.
- Shao, Z.; Song, S.; Jaiswal, S.; Shen, L.; Valstar, M.; and Gunes, H. 2021. Personality Recognition by Modelling Person-specific Cognitive Processes using Graph Representation. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM ’21, 357–366. New York, NY, USA: Association for Computing Machinery. ISBN 9781450386517.
- Song, S.; Jaiswal, S.; Sanchez, E.; Tzimiropoulos, G.; Shen, L.; and Valstar, M. 2023a. Self-Supervised Learning of Person-Specific Facial Dynamics for Automatic Personality Recognition. *IEEE Transactions on Affective Computing*, 14(1): 178–195.
- Song, S.; Jaiswal, S.; Shen, L.; and Valstar, M. 2022. Spectral Representation of Behaviour Primitives for Depression Analysis. *IEEE Transactions on Affective Computing*, 13(2): 829–844.
- Song, S.; Shao, Z.; Jaiswal, S.; Shen, L.; Valstar, M.; and Gunes, H. 2023b. Learning Person-Specific Cognition From Facial Reactions for Automatic Personality Recognition. *IEEE Transactions on Affective Computing*, 14(4): 3048–3065.
- Sonlu, S.; Doğan, Y.; Ergüzen, A. Ü.; Ünal, M. E.; Demirci, S.; Durupinar, F.; and Gündükbay, U. 2024. Towards Understanding Personality Expression via Body Motion. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 628–631. IEEE.
- Soubelet, A.; and Salthouse, T. A. 2011. Personality–cognition relations across adulthood. *Developmental psychology*, 47(2): 303.
- Spoerer, C. J.; McClure, P.; and Kriegeskorte, N. 2017. Recurrent convolutional neural networks: a better model of biological object recognition. *Frontiers in psychology*, 8: 1551.
- Suman, C.; Saha, S.; Gupta, A.; Pandey, S. K.; and Bhat-tacharyya, P. 2022. A multi-modal personality prediction system. *Knowledge-Based Systems*, 236: 107715.
- Tuber, S. 2012. *Understanding personality through projective testing*. Jason Aronson.
- VandenBos, G. R. 2007. *APA dictionary of psychology*. American Psychological Association.
- Ventura, C.; Masip, D.; and Lapedriza, A. 2017. Interpreting cnn models for apparent personality trait regression. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 55–63.

- Vinciarelli, A.; and Mohammadi, G. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing*, 5(3): 273–291.
- Wang, L.; Chen, Z.; Yu, T.; Ma, C.; Li, L.; and Liu, Y. 2022. Faceverse: a fine-grained and detail-controllable 3d face morphable model from a hybrid dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 20333–20342.
- Wei, X.-S.; Zhang, C.-L.; Zhang, H.; and Wu, J. 2018. Deep Bimodal Regression of Apparent Personality Traits from Short Video Sequences. *IEEE Transactions on Affective Computing*, 9(3): 303–315.
- Yang, C.; Xu, Y.; Shi, J.; Dai, B.; and Zhou, B. 2020. Temporal Pyramid Network for Action Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Yang, H.; Gee, J.; and Shi, J. 2024. Brain Decodes Deep Nets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 23030–23040.
- Ye, P.; Wang, T.; and Wang, F.-Y. 2018. A survey of cognitive architectures in the past 20 years. *IEEE transactions on cybernetics*, 48(12): 3280–3290.
- Yegnanarayana, B. 2009. *Artificial neural networks*. PHI Learning Pvt. Ltd.
- Zhang, L.; Peng, S.; and Winkler, S. 2022. PersEmoN: A Deep Network for Joint Analysis of Apparent Personality, Emotion and Their Relationship. *IEEE Transactions on Affective Computing*, 13(1): 298–305.
- Zhu, H.; Kong, X.; Xie, W.; Huang, X.; Shen, L.; Liu, L.; Gunes, H.; and Song, S. 2024. PerFRDiff: Personalised Weight Editing for Multiple Appropriate Facial Reaction Generation. In *ACM Multimedia 2024*.