Mapping Personality at Scale: AI-Based Inference and Spatial Analysis

Helen X. H. Bao[a*], Yiju Zhang[b], and Haijun Bao[c]

[a] *Department of Land Economy, University of Cambridge, Cambridge, UK*
[b] *College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China*
[c] *School of Spatial Planning and Design, Hangzhou City University, Hangzhou, China*

Abstract

Personality traits are foundational to the behavioural sciences, yet their empirical study has long been constrained by the "scalability crisis" of self-report questionnaires. This paper argues that the field is undergoing a paradigm shift: from viewing personality as a latent internal attribute captured via introspection, to treating it as a scalable behavioural signal inferred via artificial intelligence. We synthesize recent advances in AI-based personality inference from diverse digital traces, including linguistic nuances, mobility patterns, and visual behaviors, published since 2020. To structure this synthesis, we introduce a Domain–Scope Framework that bridges the gap between individual-level psychology and societal-level outcomes. Our study advances three core theoretical claims. First, that AI-based inference enables a "continuous psychometry" that captures personality dynamics in ecologically valid, real-world settings. Second, that the analytical value of these inferred traits is optimized through spatial and collective aggregation, offering a solution to the long-standing challenge of linking individual differences to macro-level social structures. Third, that urban and spatial systems serve as a critical "intermediate scale" for validating the predictive power of traits at a population level. This review provides a roadmap for a more geographically and socially embedded personality psychology.

1. Introduction

Personality traits shape how people perceive information, make decisions, and interact with others. They are consistently linked to behaviour across economic choice (Rao and Lakkol 2022), social interaction (Arpaci, Karatas, Kusci, and Al-Emran 2022), subjective wellbeing (Anglim et al. 2020), and political participation (Bromme, Rothmund, and Azevedo 2022). Yet despite their centrality to behavioural science, personality traits have remained difficult to measure at scale. Most empirical research continues to rely on self-reported questionnaires, which are costly to administer, static in time, and limited in coverage (Morgeson et al. 2007). As a result, personality is often treated as an individual-level construct, observed sparsely and infrequently, rather than as a population-level behavioural signal derived from individual behaviour and capable of aggregation across contexts and over time.

Recent advances in artificial intelligence and machine learning are changing this situation. A growing body of research shows that personality traits can be inferred indirectly from digital traces generated through everyday behaviour, including language use, visual content, speech patterns, mobility, and device interaction (see, for example, Obschonka et al. 2020; Peters and Matz 2024; Ren, Shen, Diao, and Xu 2021; Yang, Lau, and Abbasi 2023). These approaches do not replace psychometric theory, but they alter how personality can be

[*] Corresponding author (hxb20@cam.ac.uk).

observed. Instead of relying on explicit self-assessment, traits are inferred probabilistically from patterns of behaviour that are already produced at scale. This shift has implications that extend beyond methodological convenience. It changes the temporal resolution, spatial coverage, and analytic scope of personality research, allowing behavioural scientists to study psychological variation across populations, environments, and institutional settings.

The emergence of AI-based personality inference also raises a broader measurement question for behavioural science. What happens when personality traits are treated not only as latent psychological attributes, but also as inferable behavioural signals that can be aggregated, compared, and mapped? This question is not primarily technical. It concerns how behavioural heterogeneity is conceptualised, how individual differences scale to collective outcomes, and how psychological variation interacts with social and spatial structure. Addressing it requires synthesis across psychology, computational methods, and the social sciences.

Urban and spatial systems provide a particularly demanding context in which to examine these issues. Cities concentrate diverse populations, intensify social interaction, and generate large volumes of digital behavioural data. At the same time, many outcomes of interest in behavioural science, such as civic participation, compliance, mobility, risk response, and cooperation, are expressed in spatially structured environments (see, for example, Chang, Weng, and Wang 2021; Crown, Gheasi, and Faggian 2020; Weinschenk 2017). Personality traits influence these behaviours, yet they are rarely incorporated into spatial or policy analysis because they are not routinely measured at scale. AI-based inference methods create the possibility of bridging this gap by linking psychological traits to places rather than only to individuals.

This review examines recent advances in AI-augmented personality assessment and evaluates their implications for behavioural research at population and societal levels. We focus on studies published since 2020 that infer personality traits from digital traces using machine learning and deep learning methods. Rather than providing an exhaustive catalogue of algorithms, we synthesise this literature around three questions. First, what kinds of behavioural signals different data modalities capture, and which personality traits they infer with relative reliability. Second, how these inferred traits can be aggregated across individuals and contexts, including spatial units such as neighbourhoods, regions, or cities. Third, what methodological and ethical constraints arise when personality is inferred without self-report and used in population-level analysis.

To structure this synthesis, we introduce a domain–scope framework that links personality traits to behavioural outcomes across economic, social, and political domains, and across individual-level and societal-level analysis. This framework clarifies where personality inference has been most commonly applied, where evidence is currently thin, and where population-level inference raises new challenges. It also provides a basis for evaluating which AI-based methods are suitable for different behavioural questions, depending on data availability, scale, and risk.

Across modalities, the literature shows consistent patterns. Text-based inference captures aspects of openness, affect, and social orientation, but is sensitive to language, context, and platform norms. Visual and audio modalities often yield higher predictive accuracy for certain traits, particularly those expressed through non-verbal behaviour, but raise stronger concerns around consent and surveillance. Smartphone-derived behavioural data offer longitudinal and context-rich signals that are well suited to population analysis, yet they reflect patterns of device use that vary across groups and settings. Physiological signals provide valuable validation in controlled environments, but are not scalable for societal inference. Multimodal systems improve predictive performance by combining complementary signals, though they also increase complexity and opacity. Taking together, these methods open new avenues for studying how psychological composition relates to social dynamics, institutional performance, and environmental context. It also introduces risks. Inferred personality traits are probabilistic, culturally contingent, and sensitive. When treated as data layers rather than self-reported attributes, they can amplify bias, obscure uncertainty, or be misused for behavioural profiling.

By bringing together research on AI-based personality inference with behavioural and spatial analysis, this review advances two claims. First, personality can now be studied as a scalable behavioural signal, provided its inferential limits are made explicit. Second, spatial and societal contexts are not peripheral applications, but central test cases for evaluating the validity, usefulness, and ethics of these methods. How personality is inferred, aggregated, and interpreted at scale will shape not only future research, but also how behavioural insights are used in policy and governance.

In the sections that follow, we outline a conceptual framework for linking personality to behavioural outcomes, synthesise AI-based inference methods by data modality, and examine how inferred traits can be integrated into population-level analysis. We conclude by identifying priorities for validation, ethical governance, and interdisciplinary collaboration as personality measurement enters a new phase of behavioural science.

2. Review methods and classification framework

2.1 Literature identification and selection

This Review follows a structured literature identification and screening process guided by the PRISMA framework (Page et al. 2021). The aim was not to exhaustively catalogue all work on personality prediction, but to assemble a coherent and comparable body of recent studies that apply artificial intelligence and machine learning methods to infer personality traits from behavioural data, with relevance for population-level and spatial analysis. The review focuses primarily on studies that operationalise personality using the Big Five framework (Goldberg 1990). This reflects the dominant practice in the literature and enables comparison across studies that employ different data modalities and inference techniques. The use of a common trait taxonomy does not imply that alternative models of personality are unimportant, but it facilitates synthesis by providing a shared reference point for evaluating inference performance and behavioural associations.

We conducted a systematic search across four multidisciplinary and domain-specific databases: Web of Science, Scopus, PsycINFO, and the ACM Digital Library. These databases were selected to capture research spanning psychology, behavioural science, and computer science, reflecting the interdisciplinary nature of AI-based personality inference. Searches were restricted to titles and abstracts and focused on studies published from 2020 onwards, in order to reflect current methodological capabilities and modelling practices.

The search strategy combined terms related to personality constructs and trait models with terms related to artificial intelligence, machine learning, and digital behavioural data. Personality-related keywords included general descriptors (for example, “personality”, "psychological"), established trait frameworks (for example, "Big Five", "five factor model", "OCEAN"), and individual traits (for example, Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). Method-related terms captured a wide range of computational approaches and data sources, including artificial intelligence (AI), machine learning (ML), deep learning, neural networks, language models, social media, and multimodal data. Wildcards were used to capture plural and variant forms of key terms.

The initial search yielded 728 records. After removing duplicate entries and excluding studies published before 2020, 350 records remained for further screening. Titles and abstracts were then reviewed to assess relevance. Studies were excluded if they did not involve AI- or ML-based inference of personality traits, focused exclusively on theoretical psychology without empirical modelling, or lacked a clear behavioural data source. Following this screening stage, 95 studies were retained for full review and synthesis. Details of our search strategy, inclusion and exclusion criteria, and results in each step are outlined in Figure 1.

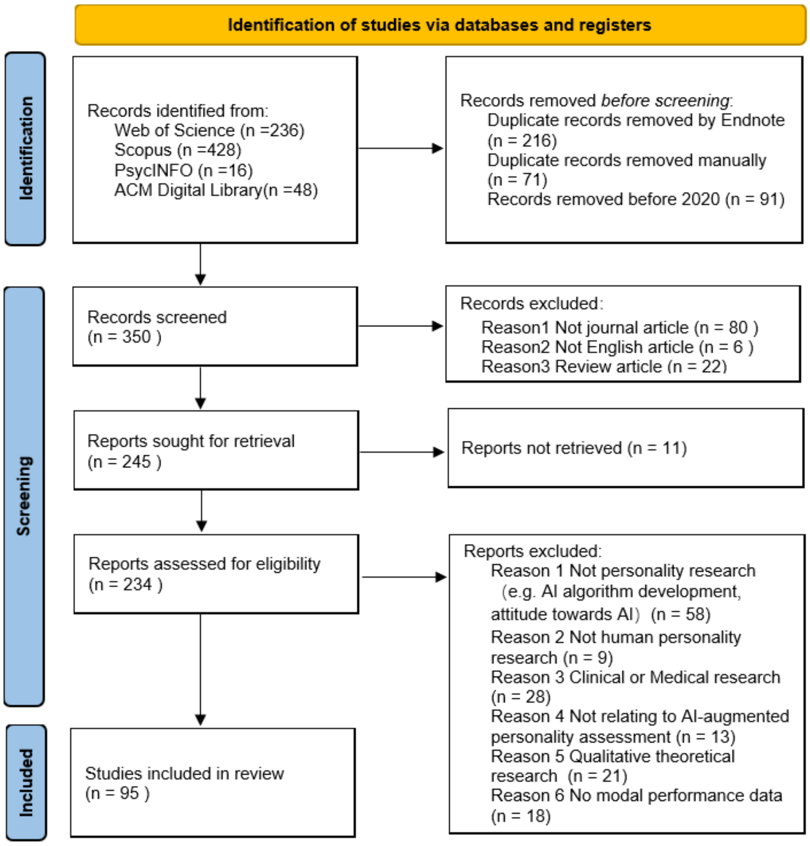


Figure 1 | PRISMA Flow Chart of Study Selection

2.2 Analytical framework for literature synthesis

The final corpus of studies spans a wide range of data modalities, including text, visual content, audio, smartphone-derived behavioural traces, physiological signals, and multimodal systems. Because the reviewed studies differ substantially in data sources, modelling approaches, and evaluation metrics, a further analytical framework was required to enable meaningful synthesis. We classified the reviewed studies using a two-dimensional analytical framework based on behavioural domain and analytical scope (see Figure 2). This framework was developed to organise evidence systematically while preserving the substantive differences between individual-level inference and societal-level outcomes, which are often conflated in studies using AI-based behavioural data.

The first dimension, behavioural domain, captures the type of behaviour or outcome examined in each study. Drawing on established divisions in behavioural and social science research, we distinguish three broad domains: economic, social, and political. The economic domain includes behaviours and outcomes related to labour market activity, consumption, productivity, and entrepreneurship. The social domain covers behaviours associated with wellbeing, compliance, cooperation, social interaction, and crime. The political domain includes attitudes, participation, trust, and institutional performance. These domains are not intended to be exhaustive, but they provide a coherent structure for grouping studies with comparable behavioural aims.

The second dimension, analytical scope, distinguishes between individual-level and societal-level analysis. Individual-level studies examine how inferred personality traits relate to differences between persons, such as variation in attitudes, preferences, or behavioural tendencies. Societal-level studies, by contrast, aggregate inferred traits across populations or spatial units, such as neighbourhoods, regions, or countries, and relate these distributions to collective outcomes. This distinction is central to the present review, as the analytical challenges and interpretive stakes differ substantially across these levels. More importantly, aggregation across individuals does not imply that societal outcomes are simple sums of individual behaviours. In many contexts, including voting, mobility, and compliance, decisions are interdependent and shaped by institutional rules, social interaction, and feedback effects. As a result, relationships observed at the individual level may not translate linearly to collective outcomes. The domain–scope framework therefore treats individual-level and societal-level analyses as analytically distinct, even when they draw on similar inference methods or data sources.

The framework was applied to all studies retained after PRISMA screening. Each study was classified according to its primary behavioural domain and analytical scope, based on the outcome variables examined rather than the data modality or modelling approach used. Studies addressing multiple domains or levels were classified according to their dominant analytical focus. This classification provided the basis for the synthesis presented in Section 3, which examines how AI-based personality inference has been applied across domains and scales, and where evidence remains uneven.

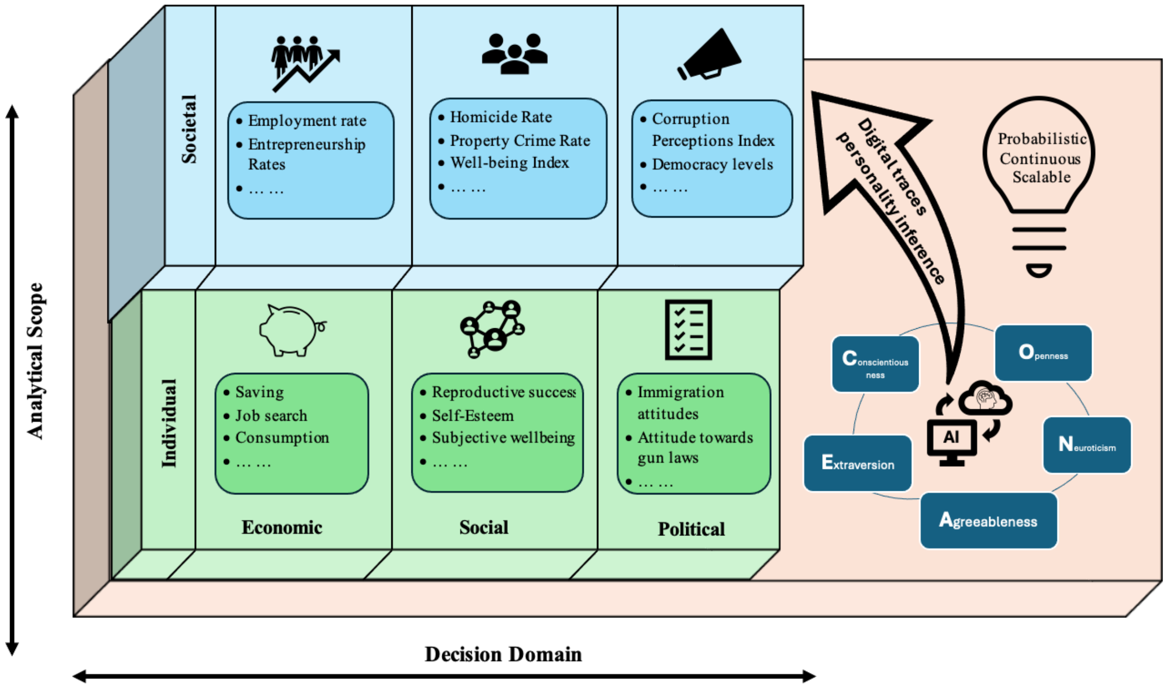


Figure 2 | Personality inference across decision domains and analytical scope.
The figure illustrates a conceptual framework linking personality traits to behavioural outcomes across decision domains (economic, social, political) and analytical scope (individual versus societal). At the individual level, personality traits are associated with behavioural outcomes such as saving, job search, subjective wellbeing, and political attitudes. At the societal level, aggregate outcomes include employment and entrepreneurship rates, crime and wellbeing indicators, and measures of institutional performance such as corruption perceptions and democratic quality. Personality traits are inferred from digital traces using AI-based methods, allowing probabilistic, continuous, and scalable estimation at the individual level and aggregation to collective patterns. The framework highlights that societal outcomes are not simple sums of individual behaviours, but emerge from interactions, contextual constraints, and the distribution of traits within populations. The Big Five model is shown to reflect its dominant use in the reviewed literature and its role in enabling comparability across studies.

3. AI-based personality inference

Using the literature identified through the PRISMA-guided search and classified according to the domain–scope framework described in Section 2, this section synthesises recent advances in AI-based personality inference. The aim is not to compare algorithms or optimise predictive performance, but to illustrate the potential of AI-enabled approaches for behavioural research by examining what kinds of behavioural signals different inference methods capture, how these signals scale across contexts and populations, and what trade-offs they entail for individual-level and societal-level analysis.

Rather than organising the review by modelling technique, we structure the synthesis around data modalities and the behavioural information they convey. This approach reflects the central premise of the paper: that the significance of AI-based personality inference lies in the new forms of behavioural evidence it enables, and in the types of questions it makes tractable across domains and levels of analysis. The subsections that follow therefore focus on the substantive implications of inference from language, visual and audio data, behavioural traces, and physiological signals, before returning to issues of aggregation, comparability, and bias.

3.1 Overview of the reviewed literature

The studies identified through the PRISMA-guided review illustrate both the rapid expansion of AI-based personality inference and the uneven development of different data modalities and measurement strategies. To provide an overview of this landscape, we summarise the distribution of studies by data modality, publication characteristics, personality frameworks, and data acquisition methods.

Data acquisition methods vary systematically across modalities and reflect clear trade-offs between accessibility and scalability (see Figure 3). Among the 95 paper reviewed, text-based inference is the most prevalent approach, accounting for nearly half of all studies (46%). Visual data represent the second largest category (18%), followed by physiological signals and multimodal systems (13%), while smartphone-derived (7%) and audio-based (4%) behavioural data remain comparatively underrepresented. This imbalance reflects differences in data accessibility, cost, and scalability across modalities, as well as the relative maturity of natural language processing compared with other sensing technologies.

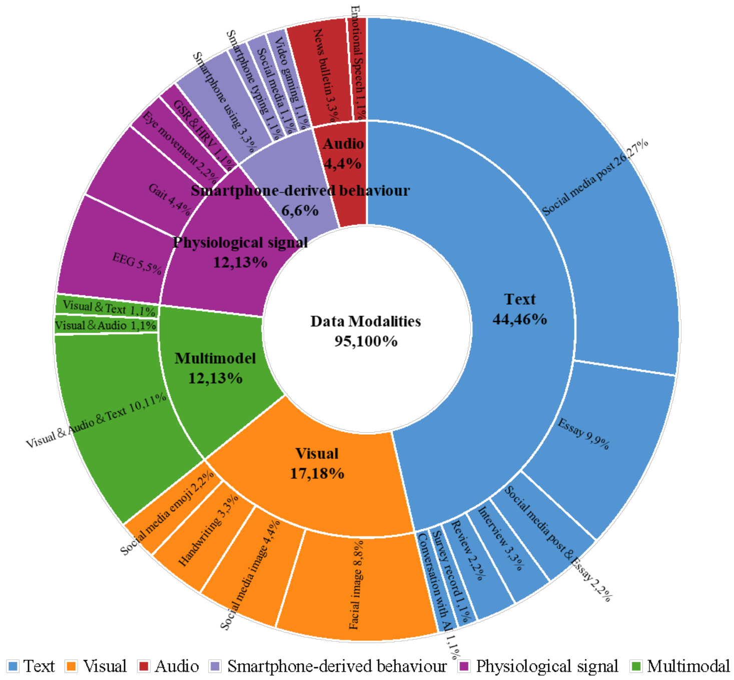


Figure 3 | Data Modality Classification

The figure shows the proportion of studies in the reviewed corpus (n = 95) that infer personality traits from different data modalities. Text-based approaches constitute the largest share of the literature, followed by visual data, physiological signals, and multimodal systems, while audio-based and smartphone-derived behavioural data are comparatively rare. Percentages indicate the share of studies using each modality as their primary data source.

Beyond modality, the reviewed studies also display clear patterns in publication timing, personality models, and data sources. The volume of publications has increased over time, with a marked acceleration after 2020, which reflects recent advances in machine learning architectures and data availability (see Figure 4a). In terms of personality frameworks, the Big Five model dominates the literature, used in approximately 86% of studies (see Figure 4b).

Other models, including MBTI, HEXACO, the Dark Triad, and Jungian typologies, appear only sporadically. This concentration around the Big Five underpins the comparative approach adopted in this review and facilitates synthesis across otherwise heterogeneous studies. Figure 4c shows that public or semi-public datasets dominate text, visual, and audio research, whereas smartphone and physiological studies rely almost exclusively on bespoke data collection. Overall, these patterns indicate that the literature is expanding rapidly and converging on a shared trait framework, yet remains concentrated in a small number of data modalities and primarily focused on individual-level inference. Current advances in AI-based personality inference are driven primarily by data availability rather than theoretical diversification.

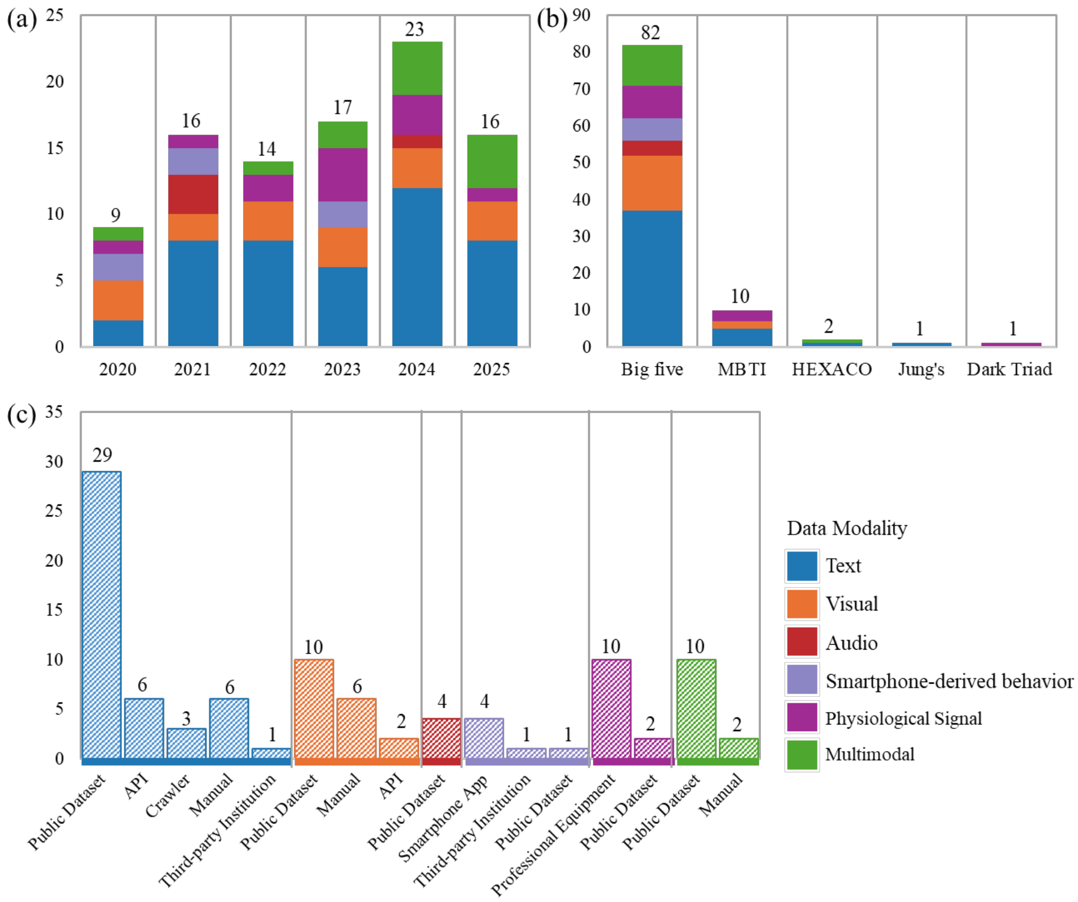


Figure 4 | Distribution of studies by data modality across (a) publication year, (b) personality models, and (c) data acquisition methods

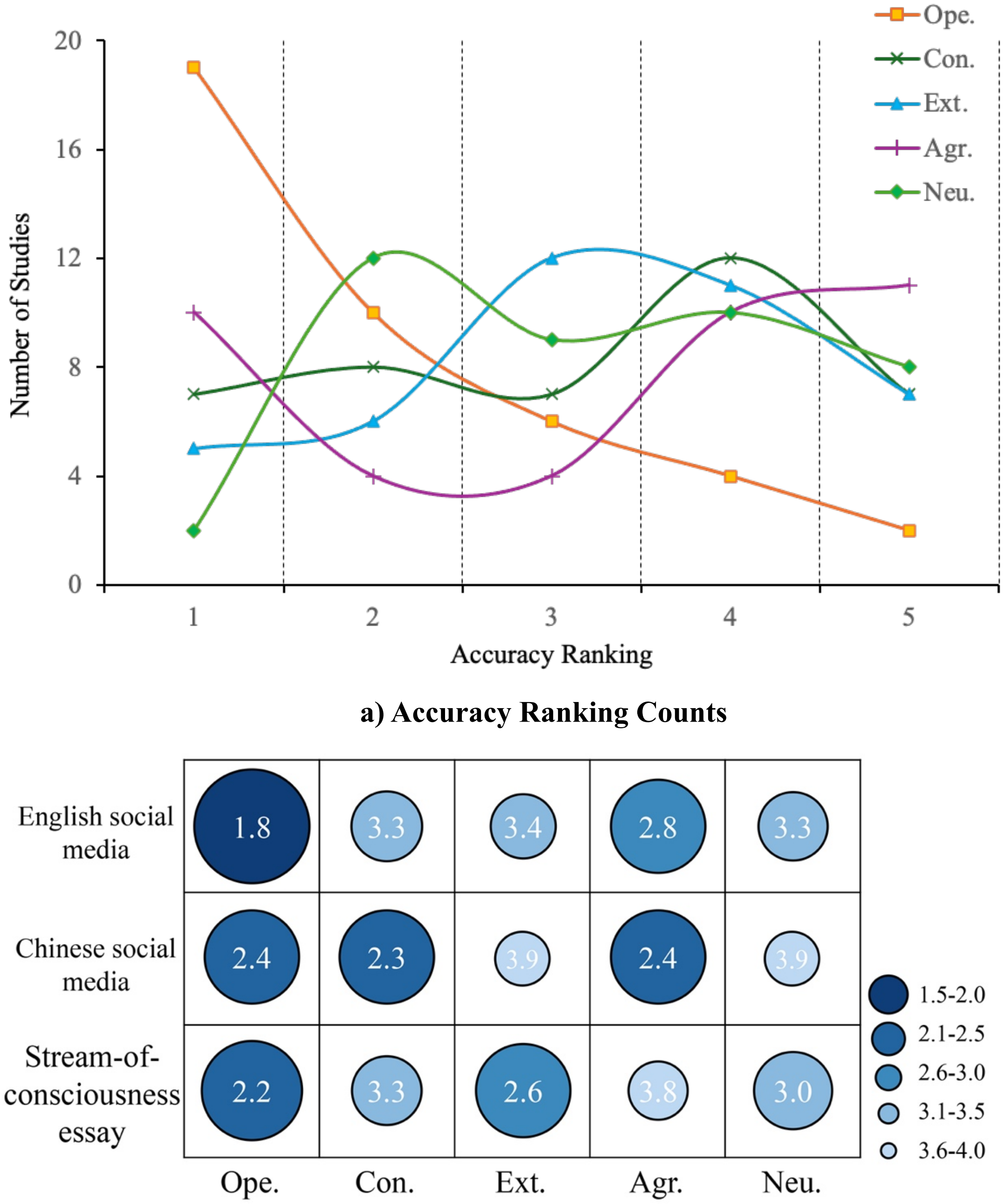
(a) gives the temporal distribution of reviewed studies published since 2020, showing a marked increase in publication volume over time. (b) shows the distribution of personality frameworks used across studies, highlighting the dominance of the Big Five model relative to alternative trait or typological approaches. (c) presents the data acquisition methods by modality, distinguishing between publicly available or benchmark datasets and bespoke data collection, including app-based and laboratory-based approaches. Together, the panels summarise key structural features of the reviewed literature.

3.2 Text-based personality prediction

Text-based approaches constitute the largest and most mature strand of AI-based personality inference in the reviewed literature (N = 44). Before examining variation across models and datasets, we first summarise trait-level patterns in predictive accuracy, then consider how these patterns vary across language contexts and text types, and finally examine how methodological choices contribute to the observed dispersion in reported performance.

Figure 5a summarises the relative predictability of the Big Five traits based on ranked accuracy within models. The X-axis of the line chart represents the ranking of prediction accuracy for different personality traits within a single model, while the Y-axis represents the number of studies. To avoid deviations in the arithmetic mean resulting from significant differences in the accuracy values across different models, we rank the accuracy values of various personality traits within each study from highest to lowest before calculating the arithmetic mean. The average predictive accuracy of Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism is 2.02, 3.10, 3.22, 3.21 and 3.24, respectively.

Across studies, openness consistently ranks among the most predictable traits. The predictive accuracy of this personality trait is ranked the first in 19 of the 41 studies included in Figure 5a. Neuroticism is ranked at or below average (≤ 3) in 27 of the 41 studies, and hence, the most difficult to predict personal trait among the Big Five. This ordering is largely stable across models, indicating that differences in trait predictability are not driven solely by specific architectures or datasets. Given the importance of language used in this type of data, we examined how trait-level predictability varies across language environments and text genres. As shown in Figure 5b, in English-language social media data, openness and agreeableness exhibit the strongest relative performance, with average rank values below 3.0. In Chinese-language social media, conscientiousness has the best average ranking score. Agreeableness remains relatively well predicted, but openness shows weaker performance relative to English contexts. In contrast, stream-of-consciousness essay data display a different pattern, with openness and extraversion ranking more favourably than in social media contexts. These differences indicate that trait predictability depends not only on the trait itself, but also on linguistic structure and cultural context.



a) Accuracy Ranking Counts

b) Average Accuracy Ranking by Types of Text Data

Figure 5 | Accuracy Ranking Comparison Based on Different Types of Text Data

(a) The X-axis of the line chart represents the ranking of prediction accuracy for different personality traits within a single model, while the Y-axis represents the number of studies. (b) Based on the three sub-data types, the accuracy ranking for different personality traits is calculated. The value represents the average rank of personality prediction accuracy, where a smaller value indicates higher accuracy for that personality prediction. Analysis includes studies using Big-Five framework (i.e., ***Ope****nness,* ***Con****scientiousness,* ***Ext****raversion,* ***Agr****eeableness and* ***Neu****roticism, or the OCEAN framework).*

Text-based personality prediction encompasses a heterogeneous set of modelling approaches that differ in both representational strategy and evaluation context. Across the reviewed studies, three broad classes of methods can be distinguished. Early work relies on static word embeddings combined with conventional machine learning classifiers (for example, Word2Vec or GloVe with support vector machines or logistic regression), while subsequent studies adopt contextualised pre-trained language models to capture semantic and syntactic variation in text (for example, ELMo, ULMFiT, and transformer-based models such as BERT and RoBERTa). More recent contributions extend these approaches through hybrid architectures or large language models, including ensemble and zero-shot systems (for example, CNN–LSTM or CNN–BiLSTM hybrids, GPT-3.5, and GPT-4). Although these methods are often evaluated using similar metrics, such as accuracy or F1 score, they differ substantially in training data, task formulation, and validation setting, which limits direct comparability across studies.

Figure 6 introduces the methodological dimension while remaining at the trait level. The centroid and dispersion ellipses show the distribution of reported accuracy values across modelling approaches for each Big Five trait. Openness and extraversion exhibit higher central tendency and tighter dispersion, indicating more stable performance across models. By contrast, conscientiousness and agreeableness display wider dispersion and lower centroids, reflecting greater sensitivity to modelling choices and data characteristics. Neuroticism occupies an intermediate position, with moderate central tendency but noticeable variance across studies. These patterns are largely consistent with the ranked results in Figure 5 while showing that differences across traits persist even when averaging across diverse modelling strategies.

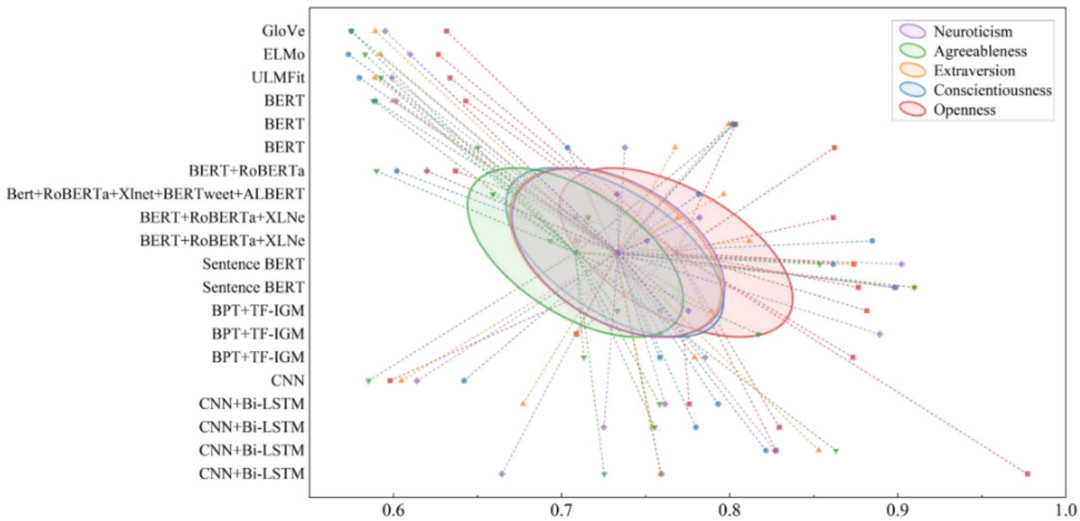


Figure 6 | Centroid Analysis of Accuracy for Big Five Traits in Text-based Studies
The centroid point at the centre of the ellipse represents the geometric centre of all data points, describing the central tendency of the data distribution. Ellipses represent a 95% confidence level region.

Figure 7 provides the most granular view by disaggregating text-based personality prediction results by individual studies, datasets, and modelling approaches. Reported accuracy values range from approximately 0.55 to above 0.90, depending on model architecture, dataset, and evaluation setting. Transformer-based models and ensemble architectures frequently achieve higher accuracy , with several studies reporting values above 0.85 on platform-specific datasets such as Instagram (Kamalesh and Bharathi 2022) and X/Twitter (Christian, Suhartono, Chowanda, and Zamli 2021; Kamalesh and Bharathi 2022). Hybrid architectures combining convolutional and recurrent components also show strong performance in some settings, though with greater variability across datasets (Aljuhani, Al-Ghamdi, Alghamdi, and Saleem 2025). The figure highlights substantial overlap in performance across model classes and demonstrates that high reported accuracy is often contingent on dataset characteristics and task formulation rather than model class alone.

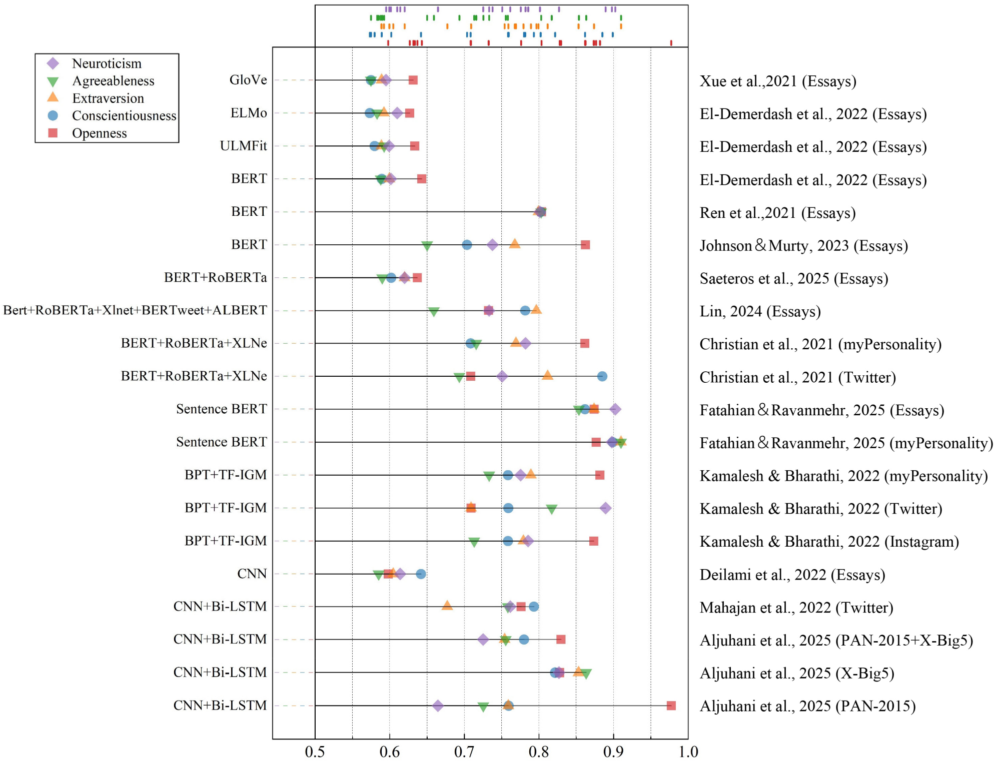


Figure 7 | Model Performance in Text-based Studies

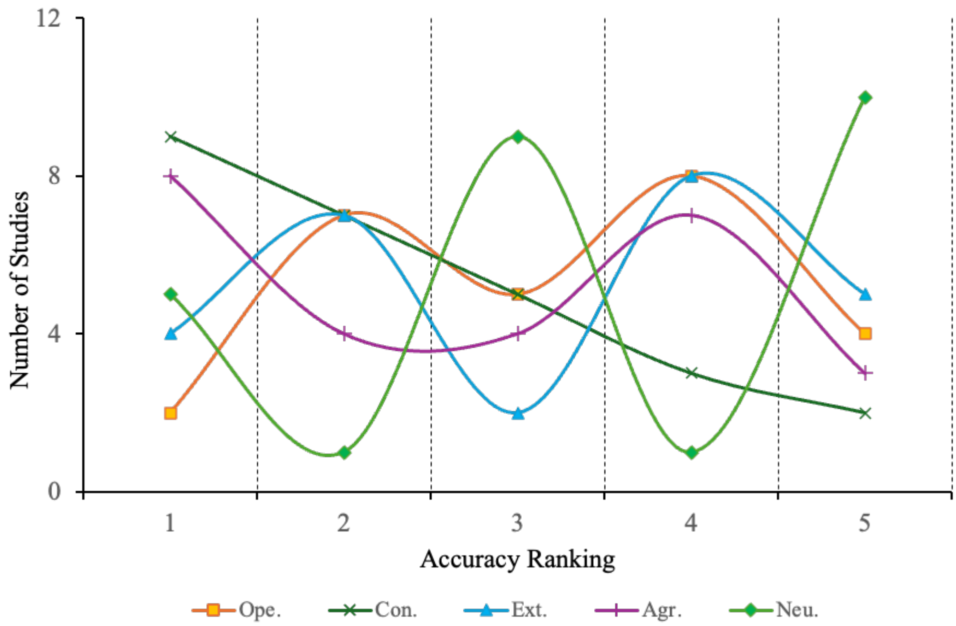
Values represent reported model performance aggregated across datasets and modelling approaches. The X-axis represents accuracy measurements for different personality prediction tasks, with 1.0 or 100% being the perfect score. The Y-axis represents various ML/DL models used (left), along with the corresponding articles and the data utilized (right). The rug at the top illustrates the distribution of prediction accuracy across different personality.

3.3 Visual Content Recognition

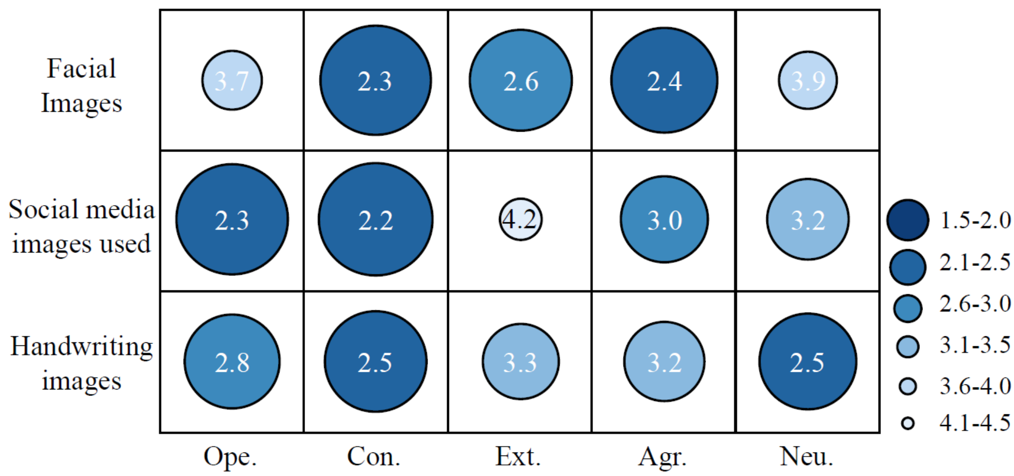
Visual-based personality prediction represents the second largest strand of AI-based personality inference in the reviewed literature. Studies in this category infer personality traits from facial images, profile photographs, videos, or other visual artefacts shared in digital environments. Compared with text-based approaches, visual inference relies on more constrained but information-dense behavioural signals. At an aggregate level, visual-based studies report uneven predictability across the Big Five traits. The average predictive accuracy of Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism is 3.19, 2.31, 3.12, 2.73, and 3.38, respectively.

As shown in Figure 8a, conscientiousness and agreeableness are most frequently identified as the most predictable traits from visual data, whereas neuroticism shows weaker and less consistent performance. Openness and Extraversion occupy an intermediate position, with moderate predictive accuracy reported in some settings but substantial variability across studies. This ordering broadly mirrors, but does not replicate, patterns observed in text-based inference, indicating partial convergence in which socially expressive traits are more readily inferred from visual cues.

Figure 8b suggests that trait predictability varies substantially across three main visual data sources and contexts identified in the literature. Static profile images, particularly those drawn from social media platforms, tend to yield stronger performance for openness and conscientiousness, whereas video-based data capture additional temporal and expressive cues that support inference of neuroticism and, in some cases, agreeableness. Controlled image datasets collected under laboratory or semi-structured conditions often report higher overall accuracy than in-the-wild social media images, suggesting that visual noise and contextual ambiguity play a significant role in shaping inference outcomes. These differences indicate that visual personality inference is highly sensitive to both the modality of visual input and the situational context in which images are produced.



a) Accuracy Ranking Counts



b) Average Accuracy Ranking by Types of Visual Data

Figure 8 | Accuracy Ranking Based on Different Types of Visual Data

(a) The X-axis of the line chart represents the ranking of prediction accuracy for different personality traits within a single model, while the Y-axis represents the number of data labels/sets. (b) Based on the three sub-data types,

the accuracy ranking for different personality traits is calculated. The value represents the average rank of personality prediction accuracy, where a smaller value indicates higher accuracy for that personality prediction.

Figure 9 summarises the distribution of reported accuracy values across modelling approaches for each Big Five trait. Deep convolutional neural networks (e.g., CNN and CNN+LSTM) dominate the literature, with newer architectures generally exhibiting higher central tendency than earlier feature-based or shallow models. Moreover, dispersion remains across traits, albeit to a less degree than the pattern identified in text-based studies (i.e., Figure 6). Extraversion and openness show relatively tighter distributions, indicating greater robustness across architectures, whereas conscientiousness and agreeableness exhibit wide dispersion, reflecting sensitivity to model design and training data. Neuroticism again occupies an intermediate position, with moderate central tendency but pronounced variability. These patterns suggest that methodological advances improve average performance without eliminating trait-specific uncertainty.

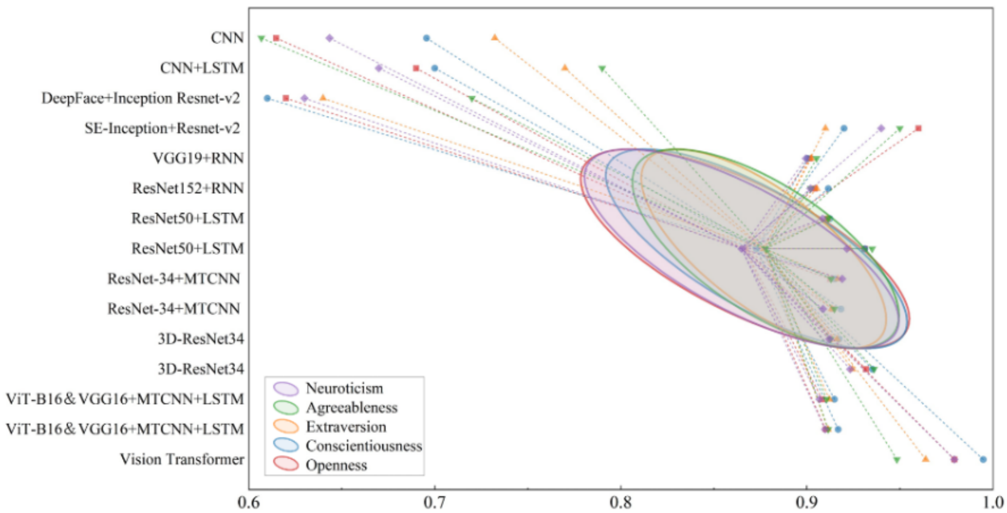


Figure 9 | Centroid Analysis of Accuracy for Big Five Traits in Visual-based Studies
The centroid point at the centre of the ellipse represents the geometric centre of all data points, describing the central tendency of the data distribution. Ellipses represent a 95% confidence level region.

Figure 10 summarises the range of modelling strategies used for visual-based personality inference across the reviewed studies. Most work relies on convolutional neural network–based architectures for feature extraction from facial images, scene images, or video frames, reflecting their capacity to capture spatially localised visual patterns relevant to behavioural expression. Residual network variants are particularly prominent, with multiple studies employing ResNet-based models of varying depth (for example, ResNet34, ResNet50, ResNet101, and ResNet152), often fine-tuned from ImageNet pre-training (e.g., Bao et al. 2024; Suman et al. 2022; Xu, Tian, Lv, and Fan 2023). Other commonly used architectures include VGG-style networks, Vision Transformers, and hybrid pipelines that combine CNNs with recurrent components to capture temporal information from video data (for instance, Bao et al. 2025; Beyan, Zunino, Shahid, and Murino 2019; Shupyliuk et al. 2025). Several studies further incorporate auxiliary techniques such as multi-task learning, attention mechanisms, and feature filtering modules to improve robustness under limited or heterogeneous data conditions (Rüegger et al. 2020; Salmani Bajestani, Khalilzadeh, Azarnoosh, and Kobravi 2024; Xu et al. 2021).

Despite their widespread use, the three most common modelling strategies, i.e., standalone CNNs, CNN–LSTM hybrids, and DeepFace–Inception–ResNetV2 pipelines, exhibit comparatively lower predictive accuracy in Figure 10, with performance clustering well below that of several less frequently used approaches. By contrast, a group of alternative or more specialised architectures achieve predictive scores exceeding 90% across all reported traits, albeit typically in more constrained evaluation settings. At an aggregate level, visual-based personality prediction therefore exhibits stronger headline performance than text-based approaches, reflecting the high information density of facial and scene-level visual cues. Figure 10 therefore highlights that gains in visual-based personality prediction are driven as much by data conditioning and experimental design as by architectural choice, contributing to the substantial dispersion in reported performance across studies.

However, the concentration of high-performing results in a subset of studies also indicates strong dependence on dataset characteristics and evaluation design. Models achieving uniformly high accuracy are typically trained and tested on curated or demographically homogeneous datasets, whereas more widely adopted architectures are applied to heterogeneous, in-the-wild visual data. Figure 10 thus shows that while visual modalities offer substantial inferential potential, reported performance is shaped as much by data conditioning and experimental context as by model architecture, contributing to the wide dispersion observed across studies.

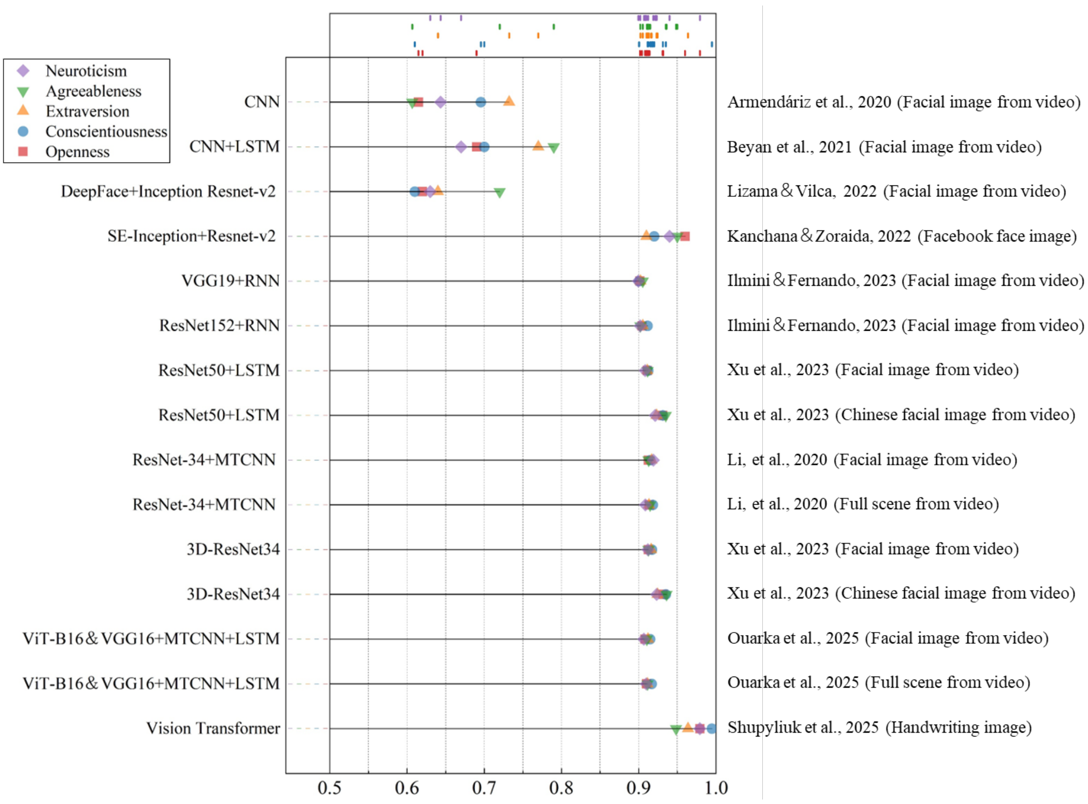


Figure 10 | Model Performance in Visual-based Studies

Values represent reported model performance aggregated across datasets and modelling approaches. The X-axis represents accuracy measurements for different personality prediction tasks, with 1.0 or 100% being the perfect score. The Y-axis represents various ML/DL models used (left), along with the corresponding articles and the data utilized (right). The rug at the top illustrates the distribution of prediction accuracy across different personality.

3.4 Audio-based personality prediction

Audio-based personality prediction remains a relatively small but distinctive strand of the AI-based personality inference literature. Studies in this category infer personality traits from vocal characteristics, speech patterns, and paralinguistic cues extracted from spoken language. Compared with text- and image-based approaches, audio-based inference draws on temporally structured behavioural signals that capture affective expression, prosody, and interactional dynamics, but it is constrained by data availability and collection costs.

Figure 11 summarises trait-level accuracy distributions across audio-based studies. Compared with text- and visual-based modalities, overall predictive performance is more modest, with most reported accuracy values clustering between approximately 0.60 and 0.80. Extraversion and neuroticism exhibit higher central tendency than agreeableness and conscientiousness, indicating that affective and expressive traits are more readily inferred from vocal signals. Openness occupies an intermediate position, with moderate average accuracy but noticeable dispersion across studies. The wide and overlapping ellipses across traits reflect both limited sample size and strong sensitivity to modelling choices, suggesting that no trait exhibits consistently robust inference across audio-based approaches.

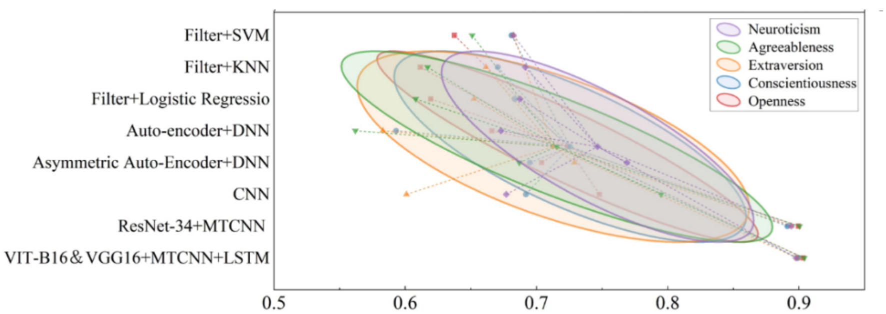


Figure 11 | Centroid Analysis of Accuracy for Big Five Traits in Audio-based Studies
The centroid point at the centre of the ellipse represents the geometric centre of all data points, describing the central tendency of the data distribution. Ellipses represent a 95% confidence level region.

Figure 12 provides a disaggregated view of model performance across the small set of audio-based studies, revealing substantial heterogeneity linked to both modelling strategy and data source. Reported accuracy varies substantially by model architecture and dataset, ranging from values below 0.65 for filter-based and conventional machine learning approaches (Liu et al. 2020; Zaferani, Teshnehlab, and Vali 2021) to values approaching or exceeding 0.90 for more complex architectures applied to curated audio data (Li, Wu, et al. 2020; Ouarka, Ait Baha, Es-Saady, and El Hajji 2024). In particular, models combining deep convolutional backbones with face detection or multimodal preprocessing pipelines achieve the highest reported accuracy (Li, Wu, et al. 2020), although these results are confined to a small number of studies using audio extracted from online video platforms. The concentration of high performance in a narrow subset of models and datasets suggests that strong results in audio-based personality prediction are highly contingent on data selection and preprocessing rather than indicative of generalisable performance.

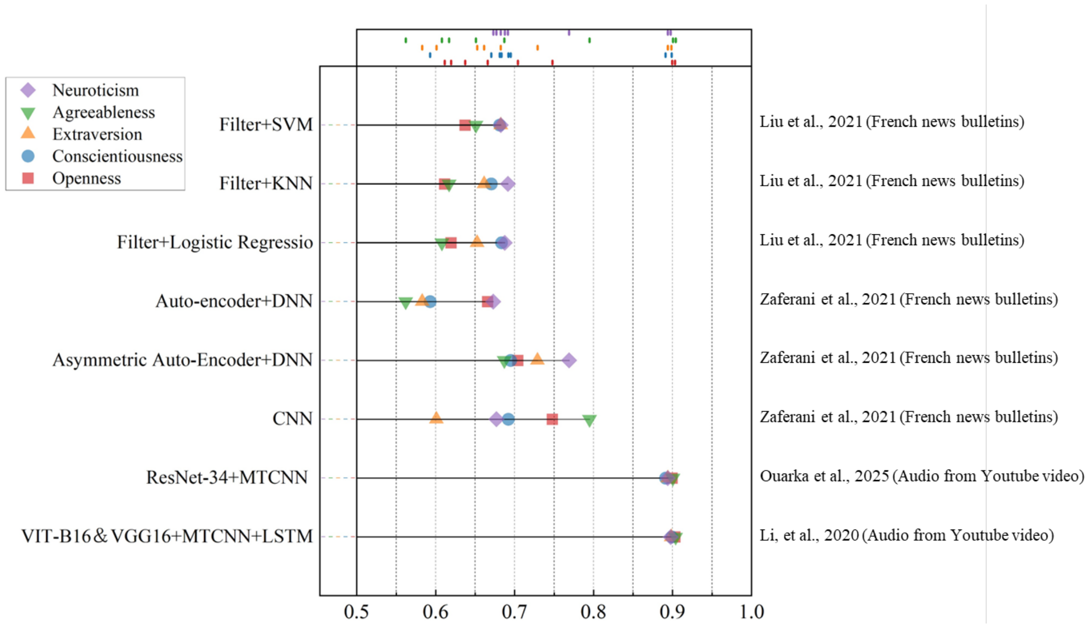


Figure 12 | Model Performance in Audio-based Studies

Values represent reported model performance aggregated across datasets and modelling approaches. The X-axis represents accuracy measurements for different personality prediction tasks, with 1.0 or 100% being the perfect score. The Y-axis represents various ML/DL models used (left), along with the corresponding articles and the data utilized (right). The rug at the top illustrates the distribution of prediction accuracy across different personality.

3.5 Smartphone-derived personality prediction

Smartphone–based personality prediction remains a small but emerging strand of the literature, with only a limited number of studies meeting the inclusion criteria. The studies reviewed in this section infer personality traits from passively collected behavioural data generated through everyday interactions with smartphones. These traces include patterns of mobility, app usage, communication frequency, screen interaction, and temporal regularity of behaviour. As summarised in Table A1 in the Appendix, the included studies differ substantially in data source, feature construction, modelling approach, and personality framework, with only one study adopting the Big Five as the primary trait model. Other studies rely on alternative or task-specific personality representations, limiting direct comparability across results.

Reported model performance varies across studies and modelling strategies. As shown in Table A1, conventional machine learning models combined with engineered behavioural features achieve moderate predictive accuracy, while deep learning approaches that capture temporal or sequential structure in behavioural traces report higher performance in some settings (Quwaider et al., 2023??). Studies incorporating longer observation windows or richer multimodal behavioural features tend to achieve stronger results than those relying on short-term or sparse traces (Quwaider et al., 2023??). However, performance metrics are not directly comparable across studies, as they differ in prediction targets, evaluation protocols, and data granularity.

Given the small number of studies and the lack of convergence on a shared personality framework, conclusions drawn from behavioural trace–based personality prediction should be treated as provisional. At present, the evidence base is insufficient to support general claims about trait predictability or comparative performance relative to other modalities. At the same time, this direction offers substantial potential. Behavioural traces are generated continuously, capture routine rather than expressive behaviour, and can be linked naturally to spatial and temporal contexts. As data availability improves and future studies adopt harmonised personality models and evaluation standards, behavioural trace–based inference may provide a powerful complement to text-, visual-, and audio-based approaches, particularly for population-level and longitudinal analysis.

3.6 Physiological Signals and Experimental Approaches

The physiological literature included in this review comprises a small set of experimental studies that infer personality traits from controlled measurements of neural, motor, ocular, or autonomic signals. As summarised in Table A2 in the Appendix, these studies draw on electroencephalography (EEG), gait dynamics captured by inertial measurement units, eye-movement data, and autonomic indicators such as galvanic skin response and heart-rate variability (Chen and Lee 2024; Li, Wu, et al. 2020; Tseng, Lin, Huang, and Lin 2023; Tsigeman et al. 2024). Data are typically collected in laboratory or semi-controlled environments using specialised instruments, and sample sizes remain modest. Although these studies differ in signal type and task design, they share a common aim of linking stable personality traits to fine-grained physiological or motor patterns under controlled conditions.

Across the included studies, model performance varies substantially by signal type and modelling strategy. EEG-based studies report relatively strong performance for Openness and weaker results for Neuroticism, a pattern supported by both individual studies and a recent meta-analysis of EEG-based personality recognition (Bhardwaj, Tomar, Sakalle, and Ibrahim 2021; Hosseini, Firoozabadi, Badie, and Azadfallah 2023; Rieck, Penava, and Buettner 2025). Gait-based studies report higher and more balanced accuracy across traits, with Agreeableness often emerging as the most predictable trait, followed by Neuroticism and Conscientiousness (Chen and Lee 2024; Ibrar et al. 2023). Accuracy increases further in settings involving abnormal or distracted gait, where movement patterns become more distinctive (Kim et al. 2025). In contrast, eye-movement-based approaches achieve low precision, typically only marginally above chance, with Extraversion showing the strongest and Conscientiousness the weakest predictive performance (Tsigeman et al. 2024; Woods, Luo, Watling, and Durant 2022). Across signal types, deep learning models generally outperform conventional classifiers, particularly when temporal dependencies are modelled explicitly (Bhardwaj, Tomar, Sakalle, and Ibrahim 2021; Sarker 2021).

Given the small number of studies and the reliance on controlled experimental settings, conclusions drawn from physiological approaches should be treated with caution. Reported performance is highly context-dependent and does not generalise readily to large-scale or population-level analysis. At the same time, physiological signals offer important opportunities. They provide high-resolution behavioural and neural measurements that can serve as validation anchors for multimodal inference, inform feature engineering in more scalable pipelines, and clarify the mechanisms through which personality traits are expressed in behaviour. As datasets expand and experimental designs become more standardised, physiological approaches may play a growing supporting role in the development and validation of AI-based personality inference systems.

3.7 Multimodal Personality Inference

Multimodal approaches combine information from multiple data streams, most commonly text, audio, and visual inputs, to mitigate the limitations of single-modality inference, such as sensitivity to noise, contextual incompleteness, or transient behavioural expression (Li, Wan, et al. 2020). The reviewed studies employ a range of fusion strategies to integrate heterogeneous signals into a unified inference pipeline. Early or feature-level fusion aggregates features across modalities prior to prediction, whereas model-level fusion concatenates intermediate representations learned separately for each modality, and late or decision-level fusion combines modality-specific predictions using voting or weighting schemes (David Curto 2021). More recent work adopts attention-based fusion mechanisms that dynamically weight modalities or features according to contextual relevance, including self-attention, cross-attention, contextual inter-modal attention, and QKV attention (Ouarka, Ait Baha, Es-Saady, and El Hajji 2024; Suman et al. 2022). Several studies further address cross-modal heterogeneity explicitly through adaptive architectures. For example, Bao et al. propose transformer-based fusion networks that learn modality correlations while suppressing redundant information (Bao et al. 2025; Bao et al. 2024), while Ryumina, Markitantov, Ryumin, and Karpov (2024) introduce gated Siamese fusion architectures that compare paired individuals during training to enhance discriminative power across modalities.

Figure 12 summarises model-level accuracy across multimodal studies and shows consistently high reported performance across a wide range of fusion mechanisms. Most multimodal models achieve accuracy values clustered above 0.85, with several exceeding 0.90 across all Big Five traits, indicating substantial gains relative to single-modality baselines reported elsewhere in the literature. Importantly, the figure shows limited separation in accuracy across different fusion strategies, including early fusion, model fusion, decision-level fusion, and attention-based mechanisms, suggesting that the use of multiple complementary data streams matters more than the specific fusion rule employed. Higher performance is most commonly reported in studies using curated benchmark datasets, such as the FIv2 corpus, and transformer-based or gated fusion architectures appear among the top-performing models (Bao et al. 2025; Bao et al. 2024; Ouarka, Ait Baha, Es-Saady, and El Hajji 2024; Ryumina, Markitantov, Ryumin, and Karpov 2024; Suman et al. 2022). At the same time, the concentration of high accuracy within a small number of datasets and evaluation settings cautions against interpreting these results as evidence of generalisable superiority. Figure 12 therefore indicates that multimodal systems offer strong inferential potential, while also highlighting the need for broader validation across contexts and populations.

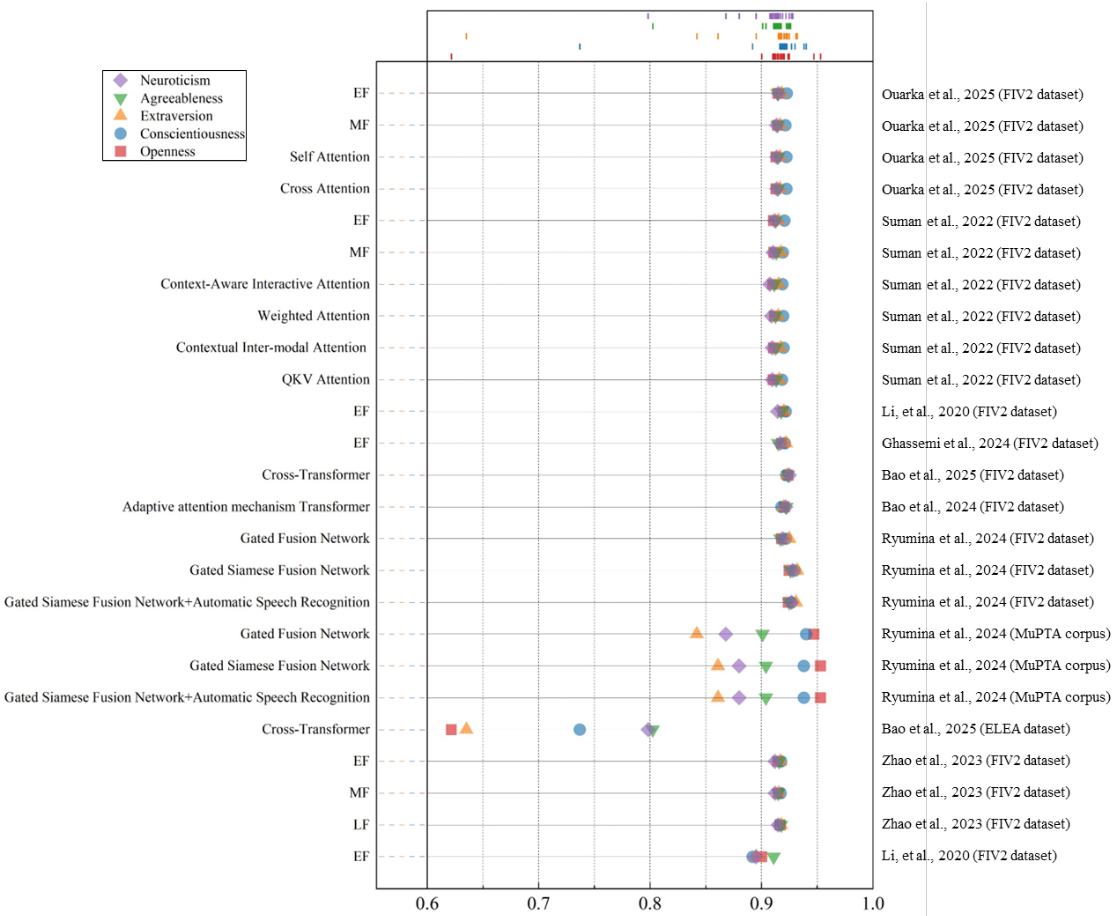


Figure 13 | Model Performance in Multimodal Studies

Values represent reported model performance aggregated across datasets and modelling approaches. The X-axis represents accuracy measurements for different personality prediction tasks, with 1.0 or 100% being the perfect score. The Y-axis represents various ML/DL models used (left), along with the corresponding articles and the data utilized (right). The rug at the top illustrates the distribution of prediction accuracy across different personality.

4. From AI-Generated Personality to Spatial Analysis

The evidence reviewed in Section 3 shows that personality traits can now be inferred probabilistically from behavioural data at scale, with varying degrees of robustness across data modalities. As these inference methods move beyond one-off assessment and become continuous, population-wide, and context-embedded, personality begins to function not only as an individual psychological attribute but also as a patterned behavioural signal. Because the contexts in which behaviour is produced, aggregated, and compared are often geographically organised, this shift creates a natural entry point for spatial analysis. Rather than treating personality as fixed or sparsely observed, AI-generated measures allow researchers to examine how psychological variation is distributed across space, how it co-varies with environmental and institutional conditions, and how these patterns evolve over time.

This section integrates them to show under what conditions inferred personality traits become spatially analysable behavioural signals. By comparing trait robustness across modalities, evaluating cost and scalability constraints, identifying methodological developments that support cross-context comparability, and situating existing applications by domain and scope, we demonstrate that spatial analysis provides a tractable and analytically justified bridge between individual-level personality inference and broader collective patterns. The proposed framework facilitates the study of how psychological tendencies are distributed, aggregated, and compared across contexts without requiring strong assumptions about social interaction or equilibrium.

4.1 Modality performance and trait robustness across inference approaches

Figure 14 summarises comparative predictive accuracy for the Big Five traits across data modalities and modelling approaches, revealing systematic differences that are analytically relevant for spatial analysis. Multimodal systems exhibit the highest and most consistent performance across all traits, with accuracy values exceeding 0.89 in every case, followed by visual and physiological signals, which show strong trait-level performance but greater variability. Text-based and audio-based approaches yield lower absolute accuracy, but display relatively stable performance across traits such as Openness, Extraversion, and Conscientiousness. Across modalities, no single trait dominates uniformly, although Openness and Conscientiousness tend to perform more robustly than Agreeableness and Neuroticism in several settings. These patterns indicate that inference quality varies jointly by trait and modality rather than being driven by either dimension alone.

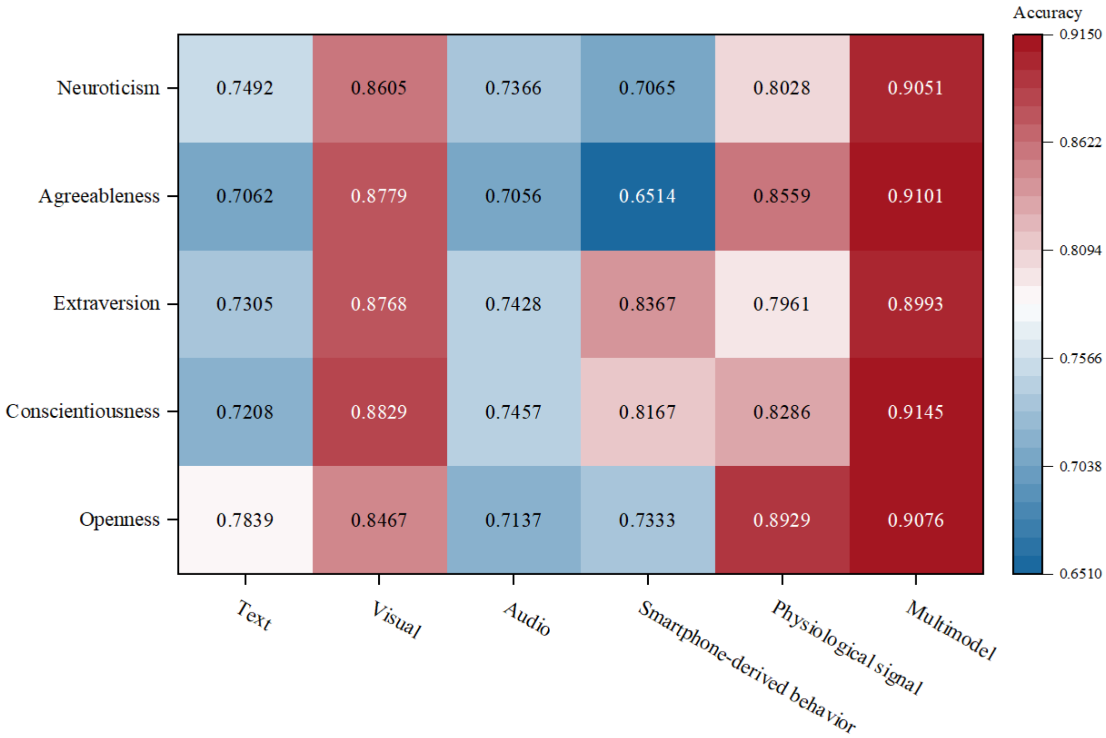


Figure 14 | Summary of Prediction Accuracy by Modality

AI-based personality prediction methods are generally divided into two categories: One treats personality prediction as a classification task, using a dichotomous approach to predict discrete category labels. Commonly used evaluation metrics include Accuracy, Precision, Recall, True Positive Rate (TPR), False Positive Rate (FPR), F1-Score, and Area Under the Curve (AUC). The other treats personality prediction as a regression task, predicting continuous personality scores. Commonly used evaluation metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2.

For spatial analysis, the implications of Figure 14 extend beyond headline accuracy. Spatial aggregation relies less on maximising individual-level prediction and more on producing consistent signals that can be compared across places and contexts. In this respect, modalities with moderate but stable performance, such as text-based and behavioural data, may be more suitable for large-scale spatial analysis than high-accuracy but less scalable approaches. Traits that demonstrate robustness across multiple modalities, particularly Openness and Extraversion, are more likely to yield interpretable spatial distributions when aggregated, whereas traits with greater sensitivity to context or measurement noise require more cautious spatial interpretation.

More broadly, the table highlights that AI-generated personality measures are inherently probabilistic and modality-dependent, which aligns with a spatial analytical perspective focused on variation rather than classification. Spatial analysis does not require traits to be inferred with uniform accuracy at the individual level. Instead, it benefits from the ability to compare relative differences, gradients, and distributions across spatial units. From this perspective, Figure 14 supports the view that AI-based personality inference is best deployed as a tool for examining patterned psychological variation across space, rather than for assigning definitive trait labels to individuals.

4.2 Cost, scalability, and data generation constraints across modalities

Table 1 summarises key differences in cost, scalability, and data collection modes across personality inference modalities, highlighting constraints that directly shape their suitability for spatial analysis. Text-based approaches combine low cost with high scalability, relying largely on passive data generation through publicly available or user-generated content. Smartphone-derived behavioural data also offer high scalability due to widespread device adoption, but entail higher costs and stronger privacy and consent requirements. Audio-based approaches occupy an intermediate position, requiring recording and processing infrastructure that limits coverage. Visual and physiological modalities impose the highest costs and lowest scalability, as they depend on specialised equipment, controlled environments, or sensitive forms of data capture.

Table 1 | Summary of Cost, Scalability and Data Collection Types.

Modality	Cost	Scalability	Data Collection Type
Text	Low (public data or API access)	High (geotagged, cross-platform text)	Passive (public or user-generated content)
Visual	Medium (camera systems, processing or public data)	Low to Medium (public capture limitations)	Passive (surveillance/capture), but sensitive
Audio	Low to Medium (recording & processing or public data)	Medium (requires audio input infrastructure)	Semi-passive (spoken input)
Smartphone-derived behaviour	Medium to High (device access, privacy consent)	High (mobile device ubiquity)	Passive (device-level tracking)
Physiological signal	High (sensors, lab environment)	Low (lab-bound or specialised settings)	Active (consented, lab-based sessions)

These contrasts have direct implications for spatial resolution and coverage. Modalities with low cost and passive data generation support fine-grained spatial aggregation across large populations and extended time periods, enabling comparative analysis across neighbourhoods, regions, or mobility-defined areas. By contrast, modalities with high cost and active data collection are more likely to yield sparse or selective spatial coverage, constraining their use to small-area studies or targeted validation exercises.

From a spatial analytical perspective, these constraints imply trade-offs between accuracy, coverage, and representativeness. High-accuracy modalities that cannot be deployed at scale may provide valuable benchmarks or ground truth, but they cannot support systematic spatial comparison. Conversely, scalable modalities may produce noisier individual-level estimates, but their broad coverage allows for aggregation and pattern detection across space. Table 1 therefore reinforces the view that spatial analysis with AI-generated personality requires selective modality choice, prioritising scalable and repeatable data sources while recognising the complementary role of more intensive measurement approaches.

4.3 Methodological innovations enabling spatial comparability

Table 2 summarises key methodological innovations in AI-based personality inference, highlighting a shift in emphasis from improving isolated prediction accuracy toward managing heterogeneity across data sources, modalities, and contexts. Across the reviewed studies, recent developments concentrate on multimodal fusion strategies, attention mechanisms, contrastive learning, and adaptive weighting schemes. These approaches aim to reconcile disparate behavioural signals, suppress redundant or noisy features, and stabilise inferred traits across environments. Rather than replacing existing models, such innovations are layered onto established architectures to improve robustness and consistency.

The relevance of these methodological advances for spatial analysis lies in their capacity to support comparability across places. Spatial analysis requires that inferred personality measures be meaningfully compared across locations that differ in behavioural baselines, data density, and contextual conditions. Fusion mechanisms and attention-based models explicitly address this challenge by dynamically weighting inputs according to relevance, thereby reducing sensitivity to modality-specific bias or context-specific noise. In this sense, methodological innovation functions less as a means of extracting new information and more as a tool for aligning heterogeneous observations into a common analytical space.

More broadly, Table 2 indicates that methodological progress in this literature is increasingly oriented toward integration rather than optimisation. This orientation aligns closely with the requirements of spatial analysis, where the goal is not to maximise prediction for individual observations, but to detect stable patterns and relationships across aggregated units. By improving the coherence and comparability of inferred traits across contexts, these innovations expand the analytical range of AI-generated personality measures, making them more suitable for spatial comparison and longitudinal analysis despite persistent uncertainty at the individual level.

Table 2 | Summary of Methodological Innovations

Dimension	Method	Contribution	Modality	Papers
Interpretability	Explainable AI (XAI) methods including integrated Gradients, Partial Dependence Plots (PDP) and Shapley Values (SHAP)	Overcoming the limitations of traditional "black-box" machine learning models by explaining prediction mechanisms and feature influences.	Text	Panfilov＆Turdakov, 2024
	Integrated Gradients	Scrutinizing NLP models' decision-making processes in personality assessment and verifing their alignment with established personality theories.	Text	Saeteros et al., 2025
	Shapley Additive exPlanations（SHAP）	Quantifying the differential impact of acoustic and linguistic features on personality trait.	Multimodal	Lukac, 2024
	Layer-wise Relevance Propagation (LRP)	Quantifying the contribution of each input feature to prediction outcomes through backward propagation and visualizes the results.	Text	Yang et al., 2022
	An interpretable machine-learning technique called permutation importance	Analyzing the unique contributions of behavioral categories and their importance in the overall model through permutation importance, and combining the significance test (PIMP algorithm) to explain the decision-making logic of the model.	Smartphone-derived behavior	Stachl et al., 2020
Bias mitigation	The cost-sensitive learning	Minimizing the model's total cost (or simply the loss) by imple- menting the unequal cost of misclassification between classes.	Text	Pradana＆Suhartono, 2025
	Synthetic minority oversampling technique (SMOTE)	Using "synthetic" instances to oversample the minority class to resolve unbalanced data.	Text	Ryan et al., 2023
	Asymmetric Auto-Encoder	Extracting features that provide adequate separation between low and high levels of the studied personality trait.	Audio	Zaferani et al., 2021
	A log-likelihood based BIRCH clustering of annotations	Reducing the unexplained variance caused by unknown differences in raters perception.	Audio	Liu et al., 2021
	Soft thresholding and a subnetwork to generates thresholds dynamically	pruning redundant information adaptively, enhancing model robustness against noisy data for more precise extraction of deep-level personality-related image features.	Visual	Xu et al., 2021
	Bell Loss function	The novel Bell Loss function is employed to address prediction bias in extreme values. When deviations are large, the gradient saturates as deviation increases, preventing excessive focus on outliers.	Multimodal	Li, et al., 2020
Data augmentation	A procedure based on the characteristics of the distribution of signs in the profiles of testees with varying degrees of severity of psychological properties is used	Increasing the trained neural network's generalizing capability and overcoming the problem of retraining, which is relevant for a small sample size.	Text	Panfilov＆Turdakov, 2024
	A transfer learning approach with Self-supervised Personality Detection Fine-Tuning (SPDFiT)	Enabling the transfer of domain-specific knowledge from similar source problem domains to enhance the target task of personality detection.	Text	Yang et al., 2023
	Time warping, time masking and frequency masking methods	Increasing data samples to cope with the limited samples of the dataset.	Audio	Zaferani et al., 2021
Multimodal fusion	A two-level regression approach for multimodal feature fusion	Enabling adaptive weighting of each feature subgroup based on predictive capability, preventing feature quantity dominance in predictions.	Multimodal	Koutsoumpis et al., 2024
	A novel Gated Siamese Fusion Network (GSFN)	Enabling the fusion of hand-crafted and deep features across three modalities (video, audio, and text).	Multimodal	Ryumina et al., 2024

4.4 Domain and scope: positioning spatial analysis between individual inference and societal outcomes

Figure 14 maps existing studies by analytical domain and scope, revealing a strong concentration of AI-based personality inference at the individual level across economic, social, and political domains. Most reviewed applications infer personality to explain or predict individual behaviour, such as decision-making, interaction, or task performance, with relatively few studies extending inference to aggregated or collective outcomes. Societal-level applications remain sparse and unevenly distributed across domains, indicating that the technical capacity to infer personality at scale has not yet been matched by systematic analytical deployment beyond the individual.

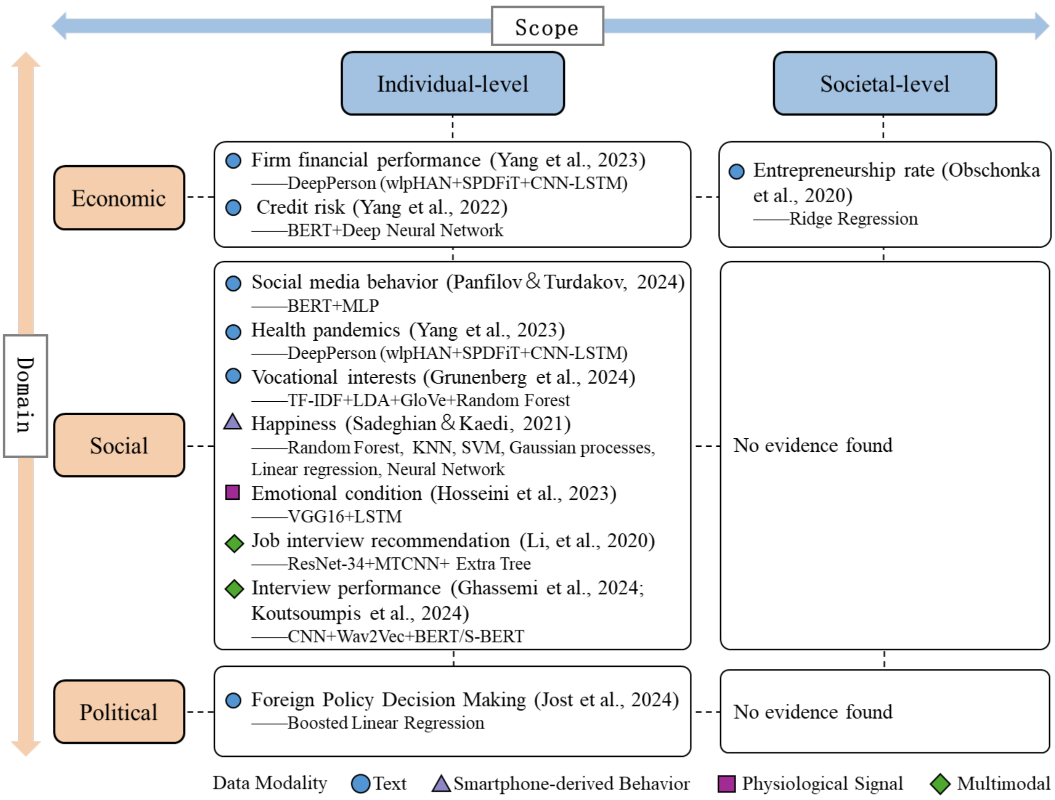


Figure 14 | AI-based personality studies by domain and scope.

This imbalance has direct implications for how AI-generated personality can be integrated into spatial analysis. Spatial aggregation offers an intermediate analytical scale that bridges individual inference and societal outcomes without requiring strong assumptions about collective decision-making or social equilibrium. By aggregating individual-level personality inferences across spatial units, researchers can examine how psychological traits are distributed across places, how they co-vary with contextual conditions, and how they relate to domain-specific outcomes at a collective level. Figure 14 suggests that such spatially aggregated analysis is currently underexplored, not because of technical infeasibility, but because existing research has remained focused on individual-level validation and performance.

From this perspective, spatial analysis emerges not as an application domain competing with individual or societal analysis, but as a methodological bridge between them. It allows AI-generated personality measures to be analysed at a scale that is sensitive to context and variation, while avoiding the complexity of modelling fully endogenous social interaction. Figure 14 thus highlights a clear opportunity for future work to extend AI-based personality inference into spatially structured analysis, using aggregation and comparison across places as a means of translating individual behavioural signals into interpretable collective patterns.

5. Research agenda and cautions for AI-based spatial personality analysis

Building on the synthesis in Section 4, we outline a focused research agenda and a set of cautions for the use of AI-based personality inference in spatial analysis. This section identifies the methodological conditions under which spatial personality analysis can be credible, the risks that arise from aggregating inferred psychological traits across space, and the limits that should guide interpretation. Together, these considerations define the boundaries within which AI-generated personality can contribute meaningfully to spatial research.

A first priority concerns methodological development tailored to spatial aggregation. Most existing studies validate personality inference at the individual level (see Figure 14), whereas spatial analysis requires robustness at the level of places. Future work should therefore prioritise place-level validation strategies, including sensitivity to spatial scale, uncertainty propagation under aggregation, and out-of-sample testing across regions or contexts. Reporting distributions, variance, and confidence intervals alongside point estimates becomes essential when inferred traits are compared across spatial units. Without such practices, spatial patterns risk reflecting artefacts of sampling density or model instability rather than substantive behavioural variation.

Relatedly, methodological choices should be aligned explicitly with spatial resolution and analytical purpose. As shown in Section 4, different data modalities support different forms of spatial analysis. Scalable text-based and behavioural trace data are more suitable for fine-grained and longitudinal spatial comparison, while high-cost modalities serve better as validation or calibration tools. Methodological innovation should therefore be evaluated not only in terms of predictive accuracy, but also in terms of how well it supports comparability across heterogeneous spatial contexts. Progress in spatial personality analysis will depend less on marginal model improvements and more on transparent alignment between data, method, and spatial scale.

A second set of concerns relates to representativeness and bias. AI-based personality inference inherits the biases of the data from which it is trained, and spatial aggregation can amplify these biases by embedding them into places. Platform-specific user populations, uneven data coverage, and differential participation across regions can lead to systematic spatial distortions in inferred trait distributions. When aggregated, such distortions risk being misinterpreted as genuine psychological differences between places rather than artefacts of data generation processes.

These risks underscore the need for careful documentation of data provenance, population coverage, and missingness when conducting spatial personality analysis. Researchers should be explicit about whose behaviour is being captured, which places are underrepresented, and how inference quality varies across space. Where possible, triangulation across data modalities or platforms can help assess robustness, although such approaches remain unevenly developed. Without explicit attention to representativeness, spatial personality maps risk reifying inferred traits as attributes of places, rather than as context-dependent behavioural signals derived from partial observation.

The third set of cautions concerns interpretation and scope. AI-generated personality measures are probabilistic inferences, not direct observations of latent traits. When aggregated spatially, they describe distributions of inferred behavioural tendencies rather than collective preferences, intentions, or values. Spatial analysis can reveal patterns, gradients, and associations, but it does not, by itself, establish causal mechanisms or explain collective outcomes. Treating spatially aggregated personality measures as explanatory variables without recognising their inferential nature risks overinterpretation.

For this reason, spatial personality analysis should be understood as a descriptive and comparative analytical approach rather than a substitute for behavioural modelling or theory-driven explanation. As emphasised throughout this review, societal-level outcomes are shaped by interaction, institutions, and feedback processes that are not captured by simple aggregation. Spatial analysis provides a tractable intermediate scale at which psychological variation can be examined without assuming independent decision-making or equilibrium behaviour. Maintaining this distinction is essential if AI-based personality inference is to be used responsibly and productively in spatial research.

Finally, ethical considerations are inseparable from the methodological and interpretive issues discussed above, particularly because AI-based personality measures are inferred without explicit self-report and are often generated from behavioural data produced for other purposes. When such inferences are aggregated spatially, ethical risks shift from individual misclassification to the attribution of psychological characteristics to places or populations. This raises concerns about consent, transparency, and the potential reification of inferred traits as stable properties of groups or locations. Addressing these concerns requires careful communication of uncertainty, clear documentation of data sources, and restraint in how spatial personality patterns are interpreted and used, rather than the assumption that technical accuracy alone resolves ethical risk.

6. Conclusions

This study was motivated by a growing disconnect between the central role of personality in behavioural science and the difficulty of observing psychological variation at scale. While personality traits are known to shape decision-making, interaction, and social outcomes, conventional measurement approaches remain costly, infrequent, and limited in coverage. Recent advances in AI and machine learning have altered this landscape by enabling the inference of personality traits from behavioural data generated in everyday contexts. This development raises a broader question addressed in this paper: under what conditions can AI-generated personality be used meaningfully beyond individual-level analysis, particularly in spatially structured settings.

To address this question, we conducted a structured review of recent studies that infer personality traits from digital traces using machine learning and deep learning methods. Using a PRISMA-guided search strategy, we synthesised evidence across data modalities, modelling approaches, and application domains. Rather than focusing on algorithmic novelty, the review classified studies by modality, trait coverage, scalability, and analytical scope, and evaluated how these dimensions interact. A domain–scope framework was used to situate existing work and to identify where inference has been concentrated and where gaps remain.

The review yields three main findings. First, AI-based personality inference is now technically feasible across multiple data modalities, with consistent patterns of trait robustness emerging across text-based, behavioural, and multimodal approaches. Second, scalability and data generation constraints shape the analytical use of inferred personality more strongly than headline accuracy, with only a subset of modalities supporting aggregation across space and time. Third, despite technical progress, most existing applications remain focused on individual-level outcomes, leaving spatially aggregated and societal-level analysis comparatively underdeveloped. Together, these findings suggest that AI-generated personality is best understood as a probabilistic behavioural signal whose value lies in aggregation, comparison, and contextual analysis rather than deterministic classification.

The contribution of this Review is threefold. Substantively, it consolidates a rapidly expanding literature on AI-based personality inference and clarifies how performance, scalability, and methodological innovation jointly determine analytical potential. Conceptually, it shows that spatial analysis provides a tractable intermediate scale between individual inference and collective outcomes, allowing psychological variation to be examined across contexts without strong assumptions about social interaction or equilibrium. Methodologically, it outlines the conditions under which spatial personality analysis can be credible, while highlighting interpretive and ethical constraints that accompany inference at scale. By reframing personality as spatially analysable behavioural data, this Review provides a foundation for future research that seeks to integrate psychological variation into spatial and contextual analysis in a principled manner.

References:

Aljuhani, N.M.; A.A. Al-Ghamdi; H.S. Alghamdi; and F. Saleem. 2025. Convolutional Bi-LSTM for Automatic Personality Recognition From Social Media Texts. *Ieee Access* **13:65582-65603**.

Anglim, J.; S. Horwood; L. Smillie; R. Marrero; and J. Wood. 2020. Predicting Psychological and Subjective Well-Being From Personality: A Meta-Analysis. *Psychological Bulletin* **146:279-323**.

Arpaci, I.; K. Karatas; I. Kusci; and M. Al-Emran. 2022. Understanding the social sustainability of the Metaverse by integrating UTAUT2 and big five personality traits: A hybrid SEM-ANN approach. *Technology in Society* **71**.

Bao, Y.; X. Liu; X. Li; Z. Wang; and Y. Qi. 2025. Weighted cross-integrated fusion network based on knowledge distillation for multi-modal personality recognition. *Applied Intelligence* **55:1-18**.

Bao, Y.; X. Liu; Y. Qi; R. Liu; and H. Li. 2024. Adaptive information fusion network for multi-modal personality recognition. *Computer Animation and Virtual Worlds* **35:e2268**.

Beyan, C.; A. Zunino; M. Shahid; and V. Murino. 2019. Personality traits classification using deep visual activity-based nonverbal features of key-dynamic images. *IEEE Transactions on Affective Computing* **12:1084-1099**.

Bhardwaj, H.; P. Tomar; A. Sakalle; and W. Ibrahim. 2021. Eeg-based personality prediction using fast fourier transform and deeplstm model. *Computational Intelligence and Neuroscience* **2021:6524858**.

Bromme, L.; T. Rothmund; and F. Azevedo. 2022. Mapping political trust and involvement in the personality space-A meta-analysis and new evidence. *Journal of Personality* **90:846-872**.

Chang, Y.B.; D.L.C. Weng; and C.H. Wang. 2021. Personality traits and the propensity to protest: a cross-national analysis. *Asian Journal of Political Science* **29:22-41**.

Chen, S.-J. and Y.-J. Lee. 2024. Classification of gait variation under mental workload in big five personalities. *Gait & Posture* **113:123-129**.

Christian, H.; D. Suhartono; A. Chowanda; and K.Z. Zamli. 2021. Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data* **8**.

Crown, D.; M. Gheasi; and A. Faggian. 2020. Interregional mobility and the personality traits of migrants. *Papers in Regional Science* **99:899-914**.

David Curto, A.C., Javier Selva, Sorina Smeureanu, Julio C. S. Jacques Junior, David Gallardo-Pujol, Georgina Guilera, David Leiva, Thomas B. Moeslund, Sergio Escalera, Cristina Palmero. 2021. Dyadformer: A multi-modal transformer for long-range modeling of dyadic interactions. *Proceedings of the IEEE/CVF international conference on computer vision*:**2177-2188**.

Goldberg, L.R. 1990. An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology* **59:1216-1229**.

Hosseini, M.S.K.; S.M. Firoozabadi; K. Badie; and P. Azadfallah. 2023. Personality-based emotion recognition using EEG signals with a CNN-LSTM network. *Brain sciences* **13:947**.

Ibrar, K.; A.M. Fayyaz; M.A. Khan; M. Alhaisoni; U. Tariq; and S. Jeon. 2023. Human Personality Assessment Based on Gait Pattern Recognition Using Smartphone Sensors. *Computer Systems Science & Engineering* **46**.

Kamalesh, M.D. and B. Bharathi. 2022. Personality prediction model for social media using machine learning Technique. *Computers & Electrical Engineering* **100**.

Kim, H.-E.; D.-H. Park; C.-H. An; M.-Y. Choi; D. Kim; and Y.-S. Hong. 2025. Real-Time Detection of Distracted Walking Using Smartphone IMU Sensors with Personalized and Emotion-Aware Modeling. *Sensors* **25:5047**.

Li, W.; C. Wu; X. Hu; J. Chen; S. Fu; F. Wang; and D. Zhang. 2020. Quantitative personality predictions from a brief EEG recording. *IEEE Transactions on Affective Computing* **13:1514-1527**.

Li, Y.; J. Wan; Q. Miao; S. Escalera; H. Fang; H. Chen; X. Qi; and G. Guo. 2020. Cr-net: A deep classification-regression network for multimodal apparent personality analysis. *International Journal of Computer Vision* **128:2763-2780**.

Liu, Z.-T.; A. Rehman; M. Wu; W.-H. Cao; and M. Hao. 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**.

Morgeson, F.P.; M.A. Campion; R.L. Dipboye; J.R. Hollenbeck; K. Murphy; and N. Schmitt. 2007. Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology* **60:683-729**.

Obschonka, M.; N. Lee; A. Rodríguez-Pose; J.C. Eichstaedt; and T. Ebert. 2020. Big data methods, social media, and the psychology of entrepreneurial regions: capturing cross-county personality traits and their impact on entrepreneurship in the USA. *Small Business Economics* **55:567-588**.

Ouarka, A.; T. Ait Baha; Y. Es-Saady; and M. El Hajji. 2024. A deep multimodal fusion method for personality traits prediction. *Multimedia Tools and Applications***:1-23**.

Page, M.J.; J.E. McKenzie; P.M. Bossuyt; I. Boutron; T.C. Hoffmann; C.D. Mulrow; L. Shamseer; J.M. Tetzlaff; E.A. Akl; S.E. Brennan; R.G. Chou; J. Glanville; J.M. Grimshaw; A. Hróbjartsson; M.M. Lalu; T.J. Li; E.W. Loder; E. Mayo-Wilson; S. McDonald; L.A. McGuinness; L.A. Stewart; J. Thomas; A.C. Tricco; V.A. Welch; P. Whiting; and D. Moher. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Journal of Clinical Epidemiology* **134:178-189**.

Peters, H. and S.C. Matz. 2024. Large language models can infer psychological dispositions of social media users. *Pnas Nexus* **3**.

Rao, A. and S. Lakkol. 2022. A review on personality models and investment decisions. *Journal of Behavioral and Experimental Finance* **35**.

Ren, Z.C.; Q. Shen; X.L. Diao; and H. Xu. 2021. A sentiment-aware deep learning approach for personality detection from text. *Information Processing & Management* **58**.

Rieck, C.; P. Penava; and R. Buettner. 2025. A Systematic Literature Review of Machine Learning-based Personality Trait Detection using Electroencephalographic Data. *Ieee Access*.

Rüegger, D.; M. Stieger; M. Nißen; M. Allemand; E. Fleisch; and T. Kowatsch. 2020. How are personality states associated with smartphone data? *European Journal of Personality* **34:687-713**.

Ryumina, E.; M. Markitantov; D. Ryumin; and A. Karpov. 2024. Gated Siamese Fusion Network based on multimodal deep and hand-crafted features for personality traits assessment. *Pattern Recognition Letters* **185:45-51**.

Salmani Bajestani, S.; M.M. Khalilzadeh; M. Azarnoosh; and H.R. Kobravi. 2024. Personality prediction via multi-task transformer architecture combined with image aesthetics. *Digital Scholarship in the Humanities* **39:836-848**.

Sarker, I.H. 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science* **2:1-20**.

Shupyliuk, M.; V. Martovytskyi; N. Bolohova; Y. Romanenkov; S. Osiievskyi; S. Liashenko; O. Nesmiian; I. Nikiforov; V. Sukhoteplyi; and Y. Lapchenkov. 2025. Devising an approach to personality identification based on handwritten text using a vision transformer.

Suman, C.; S. Saha; A. Gupta; S.K. Pandey; and P. Bhattacharyya. 2022. A multi-modal personality prediction system. *Knowledge-Based Systems* **236:107715**.

Tseng, C.-H.; H.-C.K. Lin; A.C.-W. Huang; and J.-R. Lin. 2023. Personalized programming education: Using machine learning to boost learning performance based on students' personality traits. *Cogent Education* **10:2245637**.

Tsigeman, E.; V. Zemliak; M. Likhanov; K.A. Papageorgiou; and Y. Kovas. 2024. AI can see you: Machiavellianism and extraversion are reflected in eye-movements. *Plos one* **19:e0308631**.
Weinschenk, A.C. 2017. Big Five Personality Traits, Political Participation, and Civic Engagement: Evidence from 24 Countries. *Social Science Quarterly* **98:1406-1421**.
Woods, C.; Z. Luo; D. Watling; and S. Durant. 2022. Twenty seconds of visual behaviour on social media gives insight into personality. *Scientific Reports* **12:1178**.
Xu, J.; W. Tian; G. Lv; and Y. Fan. 2023. Spatiotemporal fusion personality prediction based on visual information. *Multimedia Tools and Applications* **82:44227-44244**.
Xu, J.; W. Tian; G. Lv; S. Liu; and Y. Fan. 2021. Prediction of the big five personality traits using static facial images of college students with different academic backgrounds. *Ieee Access* **9:76822-76832**.
Yang, K.; R.Y. Lau; and A. Abbasi. 2023. Getting personal: A deep learning artifact for text-based measurement of personality. *Information Systems Research* **34:194-222**.
Zaferani, E.J.; M. Teshnehlab; and M. Vali. 2021. Automatic personality traits perception using asymmetric auto-encoder. *Ieee Access* **9:68595-68608**.