Model-based eye-tracking: a new window to understand individual differences and psychiatric disorders

Qianying Wu[1], Na Yeon Kim[1,2], Ralph Adolphs[1]

[1]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA
[2]Department of Psychology, University of California, Riverside, Riverside, CA, USA

Correspondence
Qianying Wu: qwu@caltech.edu

Competing Interests Statement

The authors claim no competing interests.

Abstract

Our eyes are constantly moving, and where we look reveals what we attend to, influences the decisions we make, and what we remember. While traditional laboratory-based eye-tracking protocols have generated a large body of findings based on well-controlled stimuli, modern methods now offer the ability to collect gaze data at much larger scale and with more naturalistic stimuli. Powerful computational tools also enable new analyses of high-dimensional data, incorporating feature annotation of the stimuli and model-based evaluation of gaze. These advances in both data collection and analysis are providing new insights into individual differences in both health and disease. Here we discuss four key approaches to modeling eye movement data: saliency-based attention phenotyping, data-driven gaze pattern identification, supervised machine learning classification, and unsupervised clustering. We highlight their advantages in psychiatry research, as they inform better understanding of visual attention, provide more fine-grained characterization of individual differences, and make more powerful clinical predictions. Finally, we address key methodological considerations in applying the methods and take stock of future opportunities on the horizon.

Introduction

We live in a visually rich and dynamic environment. How we explore this world with our constantly moving eyes shapes–and is shaped by–our memory, decisions, and actions. This reciprocal link makes eye-tracking a powerful complement to traditional behavioral or neural measures. It has the potential to provide a rich window into our minds, in both health and disease, and the advances we will review here argue for prioritizing its application in psychiatry research. Under typical conditions, our eyes shift two to five times per second, resulting in over 100,000 rapid eye movements (saccades) every day[1–4] – an enormous repository of data if it could be efficiently sampled. While gaze patterns often converge across individuals, pronounced individual differences also exist, some linked to psychiatric disorders with genetic underpinnings[5]. Over the past decade, eye-tracking has become a widely used tool in psychiatry research, advancing our understanding of clinically meaningful alterations in attention, memory, emotion, and decision-making[6–9]. The most recent advances in hardware and analysis tools now offer unprecedented opportunities to incorporate eye-tracking into clinical and research protocols.

Behind eye movements lies a remarkably intricate control system: 6 extraocular muscles - innervated by 3 cranial nerves - coordinate both voluntary and reflexive movements. Voluntary movements like saccades and smooth pursuit (cf. **Supplementary Information S1**) are driven by internal goals and motivations. These goals are processed in higher cortical regions, such as the frontal eye fields (FEF) and associated frontal and parietal areas, which in turn send motor commands to subcortical structures and, ultimately, brainstem nuclei, to initiate and coordinate movement[10] (**Figure 1a)**. Importantly, where we look is influenced not only by external features of the visual stimuli (e.g., reward value, visual saliency, novelty), but also by internal states (e.g., emotion, arousal) and enduring traits (e.g., anxiety, schizotypy, autism). Given this rich array of factors, computational models that incorporate both stimulus-driven and individual-level factors can provide valuable insight into attention mechanisms and their variations across individuals.

Visual attention, the mechanism through which we selectively allocate cognitive resources to a restricted set of visual features or spatial locations, is foundational to eye movements and nearly all psychological processes that operate on visual content. It resolves the challenge of processing multiple objects or features within a scene[11,12], by prioritizing those that are salient or relevant to behavioral goals (**Figure 1b**). Neurons involved in saccade planning respond both to the intended gaze locations as well as the resulting fixation targets, and also contribute to perceptual stability via mechanisms like saccadic suppression[13].

By capturing the spatial and temporal dynamics of gaze, eye-tracking provides a rich window into the cognitive processes that guide attention[14]. Variations in gaze patterns have been documented across psychiatric conditions, such as reduced attention to faces and biological motion in autism spectrum disorder (ASD)[5,15–17], biased attention towards threat in anxiety[18,19], and shorter fixation maintenance in children with Attention-Deficit/Hyperactivity Disorder (ADHD)[20,21]. With its non-invasive nature, ease of application, minimal reliance on language, and objective measures, modern eye-tracking holds exceptional promise for both research and clinical applications.

Despite this promise, a large portion of the eye-tracking studies to date have relied on relatively small samples, simple metrics (e.g. fixation durations, saccade lengths) and qualitative observations (e.g., gaze heatmaps, scanpaths, **Supplementary Information S1**)[22], and are often guided by hypotheses about pre-defined groups or case studies. These limit both the sensitivity to infer relevant cognitive processes, and the potential for data-driven discoveries. Over the past decade, however, advances in new computational tools have enabled more nuanced analyses of the intricate gaze - an opportunity ripe for more extensive application to psychiatric conditions. Moving beyond artificial and highly controlled settings, these methods offer better generalizability to the real world by capturing attention in naturalistic contexts - such as free exploration of natural scenes, gaze transitions in dynamic environments, and joint attention during live interactions[15,23,24]. Importantly, they deepen our understanding of the mechanisms that guide visual attention, allow for more fine-grained characterization of individual differences that are clinically relevant, and enhance the predictive and clinical utility of eye-tracking.

In this review, we provide a broad overview of recent developments in computational eye-tracking - an umbrella term for analytic approaches that leverage mathematical models and machine learning techniques to extract structure, patterns and predictive signals from eye-movement data. We introduce four categories of methods that have been successfully applied in psychiatric research: (1) **Visual saliency models**, which link gaze behavior to stimulus attributes and delineate individual phenotypes of visual preference; (2) **Data-driven gaze pattern identification,** which defines meaningful, multivariate areas of interest (AOIs) on a visual stimulus through the spatio-temporal features of eye movements; (3) **Supervised classification** techniques that use eye-tracking data to distinguish predefined clinical populations; and (4) **Unsupervised pattern discovery** methods that can identify subgroups or transdiagnostic categories based on shared gaze characteristics without prior labels. We discuss why they are helpful in advancing psychiatric research with examples. We also address key methodological considerations in how to choose and apply these methods that match different research purposes. We conclude by outlining current challenges and exploring future opportunities in the field.

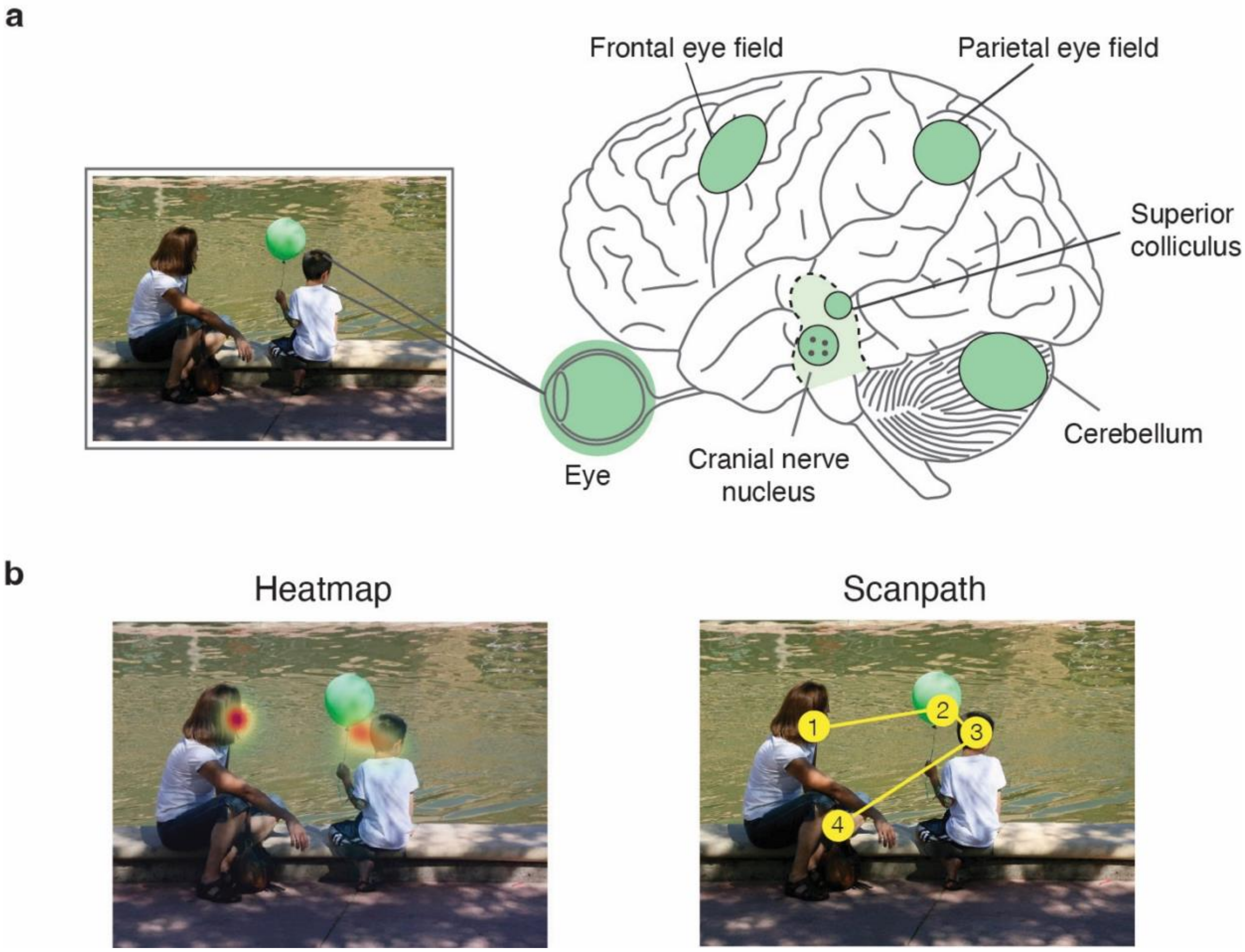


Figure 1. Studying eye movements when viewing visual stimuli. (a) Multiple brain structures are involved to initiate and control eye movements**. (b)** Eye movements, including fixations and saccades, can be visualized in the form of heatmaps or scanpaths.

Using visual saliency models to characterize attention profiles

Where we look is guided by two attention pathways: a top-down pathway driven by endogenous cognitive factors such as task and goal, and a bottom-up pathway driven by stimulus features[25]. These typically come into play concurrently during natural vision. While top-down attention dominates goal-directed search (e.g., while looking for your lost car keys), much of our daily visual experience is spontaneous, with attention captured seemingly automatically in a bottom-up manner (e.g., looking at people's faces at a party). One well-established model posits that the visual system constructs a so-called saliency map, where regions that stand out - due to color, motion, or other features - automatically attract gaze[26]. The most influential theory proposes that the features of an image are initially processed preattentively in parallel to generate the saliency map, which subsequently directs the focus of serial attention[27].

Early computational models of saliency, such as the Itti-Koch model in 1998[28], decompose an image into low-level (or pixel-level) topographic feature maps, including color, orientation, and intensity. Spatial locations that stand out from their surroundings within each feature map are combined into a master "saliency map", that partly explains eye movements driven by bottom-up visual saliency[14,28]. Following this initial model, numerous hand-crafted models have progressively improved the ability to predict fixations by incorporating high-level features defined more by their semantics than their pixelwise attributes, such as people, faces, objects, or text[29, 30–32]. Around 2015, deep neural networks (DNN) revolutionized saliency prediction by changing it to a learning problem: DNNs trained end-to-end on large sets of gaze data, were able to predict human gaze from images and could automatically incorporate higher-level features without manual specification. These DNN models achieved even higher accuracy. For example, one state-of-the-art model (DeepGaze IIE[33]) predicts human gaze on natural scene images that correlate at r = 0.82 with empirical human gaze maps, compared to r = 0.43 for the Itti-Koch model[34,35] (see the MIT/Tübingen Saliency Benchmark[36]). Beyond the prediction of gaze locations on static images, more recent work has extended gaze predictions to videos, where saliency can be driven by acoustic features and dynamic visual information (e.g., actions)[37,38]. Further extensions of these models predict scanpaths (i.e., both the direction and timing of saccades) on images by imposing other biologically- and cognitively-inspired constraints (e.g., center fixation bias, inhibition of return, evidence accumulation[39–43]) on top of the visual saliency map.

Modern saliency models now offer predictively accurate and quantitative tools to assess individual attention patterns (**Figure 2a**). Because most models are pre-trained on non-clinical samples[44–46], comparing model predictions to actual fixations can reveal which features are over- and under-prioritized by different clinical groups[47–49]. For example, one study in schizophrenia (SZ) found that high-level saliency models had better predictions for control (compared to SZ) participants' fixations on natural scenes , yet the low-level saliency model had better predictions for the SZ (compared to control) group's gaze data, suggesting a bottom-up bias in SZ patients[47]. In video watching protocols, similar dissociations were observed in ADHD: while high-level visual saliency predictions were less aligned with ADHD fixations compared to controls, low-level saliency predictions remained similar between the groups[49]. This provided evidence for the hypothesis that atypical gaze patterns in ADHD are mainly caused by reduced engagement with semantically meaningful features rather than increased sensitivity to low-level salience.

A critical extension to the diagnosis prediction of saliency models is a further examination of the individual features that drive attention (**Figure 2b**). A landmark study implemented a three-layer saliency model to examine visual attention in autistic adults during free-viewing of 700 complex natural scenes[15]. The images were comprehensively annotated at the pixel-level (3 features: color, intensity, orientation), object-level (5 features: size, complexity, convexity, solidity, eccentricity), and semantic-level (12 features: face, emotion, touched, gazed, motion, sound, smell, taste, touch, text, watchability, operability). Using a linear support vector machine (SVM) classifier, the authors evaluated the weights of each feature in predicting human fixations, and found that individuals with autism showed higher reliance on pixel-level features and reduced weighting of socially relevant semantic features[15]. This model-based approach effectively extends a particular visual feature of interest to a broader set of features, providing a multi-dimensional profile of attention biases, thus allowing researchers to test multiple hypotheses in a single experiment. While earlier studies constructed feature maps from manual annotation[15,50], more recent work leverages automated feature extraction through computer vision and large language models. Numerous open-source, validated, and user-friendly software packages have been in use for this purpose, offering automated body parts segmentation[51], facial landmark detection[52], and semantic meaning classification[16]. The field is evolving rapidly to address several remaining challenges (e.g., object tracking[53], action understanding[54], emotion recognition[55]). These tools makes it possible, with relatively little effort, to study individual differences in more naturalistic environments (e.g., virtual reality, real life) that contain richer features (e.g., fine-grained segmentation of face and body, more types of semantic features)[16,52,56–59]. Ultimately, the availability of automated annotation for large sets of features present in naturalistic stimuli could provide us with a much richer characterization of eye movements in psychiatric diseases, shedding light on more specific cognitive dysfunction.

Recent research examining the covariance structure among features that drive attention has begun to elucidate more specific latent factors[50,60–62]. For instance, individuals who tend to look at faces also show increased attention to motion, but less to bodies and text[50,62]. These studies may reveal common causes for seemingly different domains of behaviors, or different causes for seemingly similar behaviors: while both eye-to-mouth preference and face-to-object preference are considered to be measures of social attention and are linked to autism, a recent study reported weak correlation between them, suggesting that they may reflect dissociable underlying processes[61]; whereas another study showed eye-to-mouth preference co-varied with the upper-to-lower position preference when fixating on objects, suggesting a potential domain-general mechanism for processing the height of fixations. Characterizing associations of attention preference among a variety of features could also inform us about possible linkages in their genetic basis, which could ultimately suggest meaningful genetically-based endophenotypes related to psychiatric symptoms.

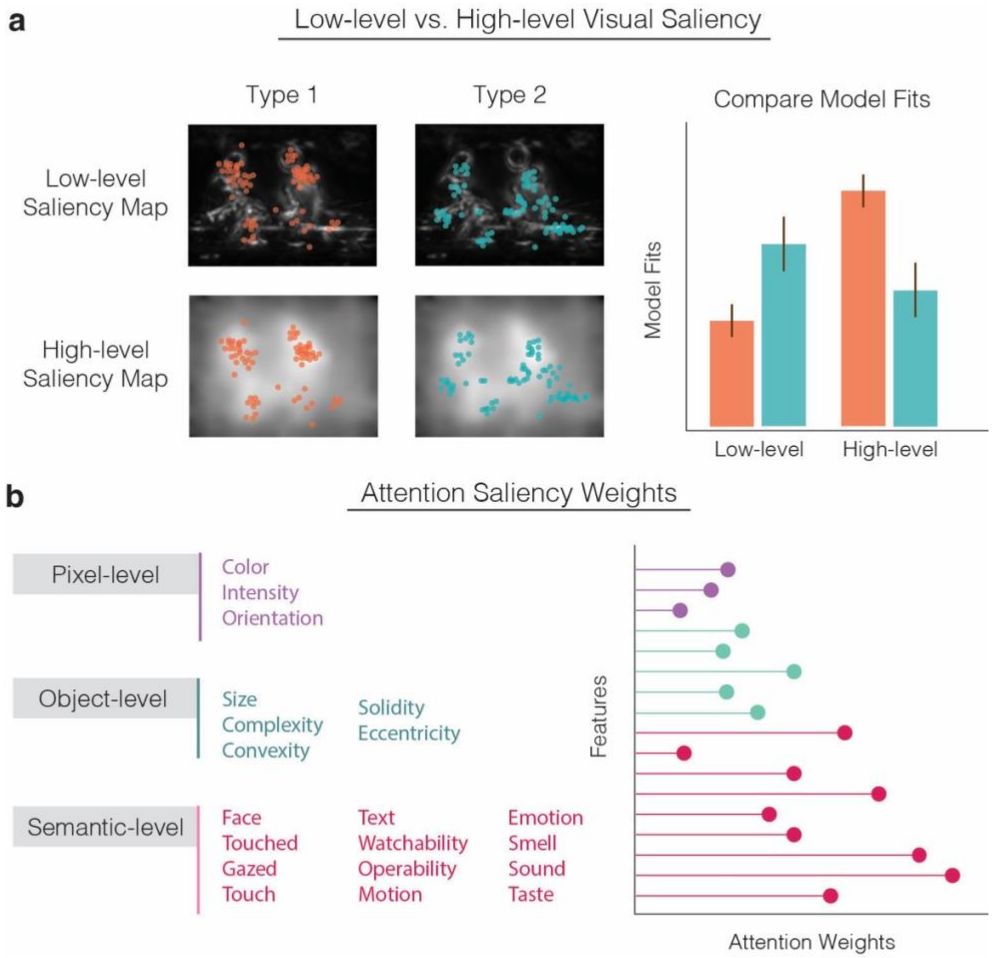


Figure 2. Using visual saliency models to understand individual differences in attention. (a) Individual gaze patterns coming from two predefined types of participants (Type 1, Type 2) are analyzed by comparing them to visual saliency maps of the stimuli. For each image (same as in Fig 1), both low-level saliency (based on features like color, intensity, and orientation) and high-level saliency (highlighting semantically meaningful elements such as people and objects) are computed. Model fit scores quantify how strongly each individual's gaze aligns with different saliency types, revealing differences in visual preferences. **(b)** Quantifying individual attention saliency weights. In one study, attention weights were derived for features across three levels: pixel-level (e.g., color, intensity), object-level (e.g., size, complexity), and semantic-level (e.g., faces, text, watchability). Weights for all the features comprise a computational phenotype for an individual's visual attention style[15].

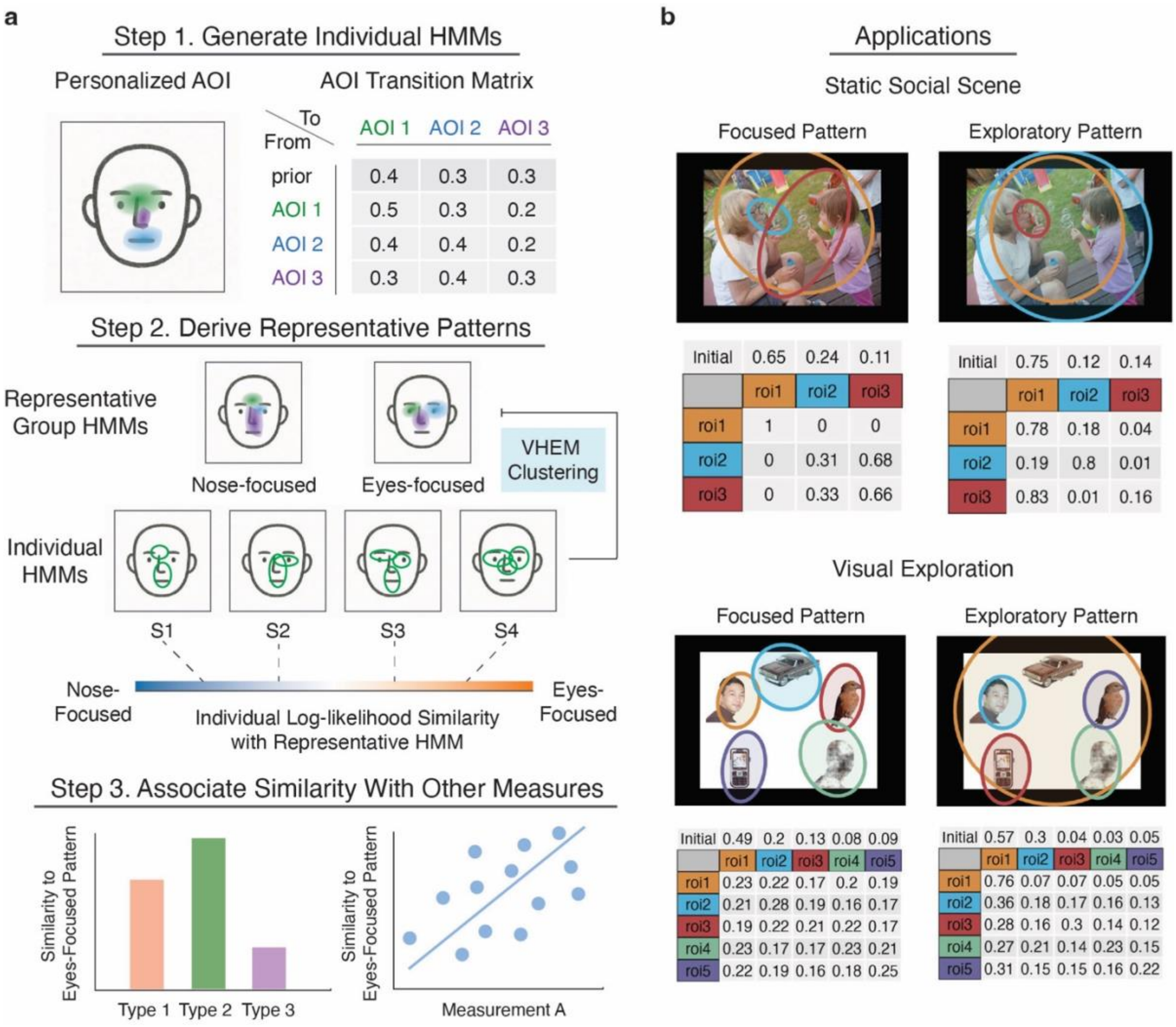
Data-driven gaze pattern identification

In a traditional protocol, researchers define a limited set of AOIs *a priori*, and then analyze metrics such as dwell time or fixation latency with respect to those AOIs. This approach has two key limitations: AOI definitions are often subjective, and informative gaze outside the AOIs may be overlooked[63,64]. For instance, when defining a “mouth” AOI on a face, one might restrict it to the lips, missing peri-oral regions that could carry relevant visual information. The many degrees of freedom for choosing and defining AOIs have been suggested to lead to inconsistent findings and decreased replicability across studies[65]. On the other hand, visualizations like heatmaps or scanpaths can highlight unexpected areas of gaze that might otherwise be missed. Yet they do not provide new AOIs – a challenge addressed by new data driven approaches that can discover interpretable gaze patterns as well as new AOIs. These algorithms identify areas that are co-fixated at adjacent times, or gaze transition sequences that are commonly observed across individuals and stimuli. Those statistical dependencies can provide the location, shape and border of a multivariate gaze pattern - an aggregate of the fixations distributed over several discrete areas. The resulting multivariate gaze patterns can then be used to define new, statistically grounded AOIs, inform hypotheses about attention priorities across populations, and eventually guide future confirmatory studies.

Multivariate gaze patterns can be identified through dimensionality reduction and data-driven clustering techniques[63,66,67]. For instance, principal component analysis (PCA) has been used to summarize complex fixation distributions into a small set of orthogonal principal components (PCs) that capture major attentional variations across individuals[63,66]. In a study that performed PCA on gaze heatmaps during face viewing, five PCs were identified that explained 82% of the total variance of fixation distributions on faces across all the participants[63]. These PCs reflected individual differences in the tendencies to fixate on the eyes relative to the mouth, and to fixate on the forehead relative to the chin[63]. Clustering methods such as K-means have also been used to derive gaze-based AOIs. Instead of relying on predefined AOIs, one study identified clusters of high-density gaze points across participants, which served to partition the image into statistically defined gaze “hot spots”[64]. Counting fixations within each data-driven AOI can then generate a lower-dimensional representation of gaze behavior for subsequent modeling of individual differences[64]. This approach enables flexible application and generalizes to diverse contexts regardless of the stimulus.

Another prominent method for finding latent spatiotemporal patterns in eye movement data is the Eye Movement Hidden Markov Model (EMHMM)[68–71]. HMM is a statistical model that describes and predicts a sequence of events. It infers the probabilistic distributions of several underlying *states* that are not directly observable (hidden) based on *observations* that are dependent on the states. In the context of eye-tracking analysis, EMHMM summarizes one's gaze locations over time ('observations') as a collection of individual-specific AOIs ('states'), as well as the transition probabilities among the states. EMHMMs first fit individual gaze data with HMMs (resulting in several AOIs for the fixations and a transition matrix describing saccades) and then cluster the individual-level results into a small number of commonly observed gaze patterns shared among participants, which are referred to as the representative group-level HMMs [72]. Individual HMMs can then be compared to the group representatives with a similarity score (per group HMM) as a continuous quantitative measure of eye movement style. Finally, depending on the research question, one can investigate the association between these HMM-derived similarity measures and other measures of individual differences, such as questionnaire-based scores (**Figure 3a**). Note that EMHMMs assume eye movements to be Markovian processes (current fixation only depends on the previous one), which simplifies the dependencies between saccades (e.g., inhibition of return and facilitation of return). Nevertheless, they are richer models than most other AOI identification methods, and provide meaningful information about the temporal dynamics of visual exploration patterns, in addition to spatial gaze locations.

The EMHMM method has been applied across various tasks and populations[68,69,71,73]. During face recognition, younger adults tend to look at faces with an analytic pattern that features frequent transitions between the eyes, while older adults, particularly those with lower executive functioning, tend to adopt a holistic pattern that starts from the nose/mouth region (i.e., high prior probability) and stays around there (i.e., high probability of transitions to the same AOI)[71,73]. The EMHMM framework has also been applied to children with autism across three social tasks (**Figure 3b**): static scene viewing, visual exploration, and activity monitoring[24]. Across all tasks, autistic children exhibited more exploratory gaze patterns than controls, characterized by broader face AOIs without eye-region prioritization, and increased attention to non-social regions. These findings illustrate how EMHMMs can uncover latent attentional styles and link them to individual traits or clinical characteristics.



To / From	AOI 1	AOI 2	AOI 3
prior	0.4	0.3	0.3
AOI 1	0.5	0.3	0.2
AOI 2	0.4	0.4	0.2
AOI 3	0.3	0.4	0.3

Static Social Scene — Focused Pattern

Initial	0.65	0.24	0.11
	roi1	roi2	roi3
roi1	1	0	0
roi2	0	0.31	0.68
roi3	0	0.33	0.66

Static Social Scene — Exploratory Pattern

Initial	0.75	0.12	0.14
	roi1	roi2	roi3
roi1	0.78	0.18	0.04
roi2	0.19	0.8	0.01
roi3	0.83	0.01	0.16

Visual Exploration — Focused Pattern

Initial	0.49	0.2	0.13	0.08	0.09
	roi1	roi2	roi3	roi4	roi5
roi1	0.23	0.22	0.17	0.2	0.19
roi2	0.21	0.28	0.19	0.16	0.17
roi3	0.19	0.22	0.21	0.22	0.17
roi4	0.23	0.17	0.17	0.23	0.21
roi5	0.22	0.19	0.16	0.18	0.25

Visual Exploration — Exploratory Pattern

Initial	0.57	0.3	0.04	0.03	0.05
	roi1	roi2	roi3	roi4	roi5
roi1	0.76	0.07	0.07	0.05	0.05
roi2	0.36	0.18	0.17	0.16	0.13
roi3	0.28	0.16	0.3	0.14	0.12
roi4	0.27	0.21	0.14	0.23	0.15
roi5	0.31	0.15	0.15	0.16	0.22

Figure 3. Data-driven AOI identification through Eye Movement Hidden Markov Models (EMHMM). (a) EMHMM procedures. Step 1: generate individual HMMs, including estimation of personalized AOIs (each color represents an AOI) and transition probabilities within and between them. Step 2: derive representative group HMMs (AOIs and their transitions) based on individual HMMs, using the variational hierarchical expectation maximization algorithm (VHEM). In the face stimuli example, two representative patterns can be nose-focused and eyes-focused patterns. The similarity between individual HMMs and representative HMMs is calculated as the log-likelihood of the representative pattern being generated by the individual. Step 3: associate similarity scores with other measures of interest. The similarity to eyes-focused pattern (relative to nose-focused pattern) becomes a behavioral phenotype, which can be compared across predefined types of participants, or correlated with measurement of interest. **(b)** Beyond face stimuli, EMHMM can be applied to a variety of images, such as static social scenes and visual exploration stimuli. Adapted from [24,70].

Supervised classification and prediction of clinical groups

The diagnosis of psychiatric disorders relies heavily on clinical observations and patient self-reports, which can be expensive, time-consuming, and subjective[74–76]. Here, eye-tracking can offer a rich array of objective and efficient measures that complement the assessment of mental health. Although specific eye-tracking metrics have been identified as particularly informative for some disorders (e.g., gaze at faces in the case of autism[77–80]; abnormal smooth pursuit in the case of schizophrenia[81,82]), the heterogeneity of psychiatric disorders generally precludes any eye movement measure in isolation from serving as a reliable clinical biomarker[83]. Machine learning (ML) models address this challenge by weighing and combining multiple features that result in more accurate and robust prediction.

In a supervised ML pipeline, researchers extract features from eye-tracking data as input (e.g., fixations within specific AOIs) and combine these features (linearly or nonlinearly) to train a model that distinguishes between groups. To avoid overfitting, the model is typically trained on only a subset of the data, and then tested on a held-out portion; doing so many times produces a distribution of model accuracies. It is also important to further evaluate the model's generalizability (without any re-training) to entirely new datasets (for an overview of supervised machine learning, see [84]). A plethora of models have achieved good accuracy in predicting various psychiatric disorders (e.g., ASD, ADHD, schizophrenia, depression, and OCD) [64,79,85–90,90,91]. A recent meta-analysis of 24 studies (2016-2021) reported a pooled classification accuracy of 81%, specificity of 79%, and sensitivity of 84% when classifying ASD from typically developed individuals[92]. Note, however, that many studies have not extensively tested generalizability to broader samples (e.g., other cultures, broader age ranges), and also have typically compared one clinical group to controls, rather than assessed specificity among multiple psychiatric conditions.

The choice of ML model and feature selection highly depend on sample size, computing resource, and desired interpretability (**Figure 4a**). Classic ML algorithms such as regularized linear regression, support vector machine (SVM), random forest, and boosting classifiers are commonly used in combination with hand-crafted features to achieve high interpretability. These models can be trained using a personal computer with several hundred samples. A variety of eye-tracking features have been tested on such models, including (1) AOI-based fixations (e.g., time spent looking at faces, histograms of fixation frequencies among AOI partitions)[64,85,87,93], (2) oculomotor features (e.g., fixation duration, saccade amplitude, saccade velocity, variation of fixation duration)[86,87,93,94], (3) saliency-based attention bias estimations[15,95,96], and (4) gaze transition patterns (e.g., transition probabilities assessed by HMM)[91,97]. Comparing the contribution of the features (i.e., calculate discriminability scores[98,99], conduct feature selection[100,101], check coefficients, or calculate feature importance in prediction[102,103]) can suggest the most informative metric or test paradigm that differentiates the clinical group.

A potentially powerful alternative approach is to use deep neural networks (DNN), which can learn gaze patterns from raw or lightly processed eye-tracking data with minimal manual feature engineering. They do this by fitting highly complicated nonlinear functions to predict diagnostic categories from all available feature information. While they capture subtle gaze dynamics that may be lost in traditional feature extraction and yield better predictive performance, they are prone to overfitting, require larger datasets (often thousands of trials or subjects), and demand significant computational resources.

Among DNNs, convolutional neural networks (CNNs) are widely used for image-based inputs. CNNs are feed-forward networks that use convolutional layers to learn spatial features of the input images at different scales, progressing from low-level patterns near the input layers, to high-level semantic features in deeper layers. In eye-tracking research, CNNs have been applied to extract spatial gaze patterns across scales from gaze heatmaps or scanpath (in the form of 2D image input). These gaze patterns are then weighted in the final CNN layer to predict psychiatric conditions, which often outperform classic ML methods[91,104–107]. CNN can also identify novel image features that differentiate gaze patterns between target groups[58,108]. In a recent study, CNNs were trained on natural images to predict differences in fixation maps (of these images) between ASD and control groups[108]. Learned representations of each image captured where (and on which image features) participants from each group tend to fixate more or less. The latent representations of each individual's fixated image pixels were used as input to an SVM classifier, which successfully classified ASD participants with an accuracy of 92%[108].

While images of gaze heatmaps or scanpaths capture spatial patterns, they collapse temporal information. To model temporal transitions of eye movements, recurrent neural networks (RNNs) can be helpful. RNNs (and their extensions such as LSTMs and Transformers) process sequential data, such as raw gaze coordinates or fixation sequences, by maintaining a memory of past inputs through recurrent connections. This allows RNNs to learn temporal dependencies in gaze trajectories, distinguishing, for example, between repetitive scanning and more exploratory gaze behavior. Compared to HMMs, which as we reviewed above also model temporal information, RNNs are better at modeling complex, nonlinear dependencies of gaze trajectories over longer time scales, yet their results are often less interpretable. In practice, RNNs are also often used for classifying participants into groups[88,90,109,110], and can be combined with CNNs to capture both the spatial and temporal features of eye movements[111–113]. Beyond CNNs and RNNs, more efficient neural network architectures and novel optimization mechanisms are developed every day. Researchers who want to implement ML-based classifications should closely monitor relevant benchmarks and the state-of-the-art models in the field.

Unsupervised pattern discovery for eye movement-based subtyping

While supervised classification identifies differences between predefined categories, unsupervised clustering aims to uncover natural groupings within or across groups without prior labels. This holds out the promise of discovering new subtypes within a disorder, overlap between disorders, or even entirely new proposals for diagnostic categories as such. Clustering begins by quantifying pairwise similarity between gaze patterns. Similarity may be based on selected features (e.g., reduced face fixations or short fixation durations) or on the overall gaze trajectory (e.g., looking at similar regions over time). Once the pairwise similarities are calculated, clustering algorithms (e.g., hierarchical clustering, K-means clustering) can be applied to detect groups of individuals that share similar eye movement profiles (**Figure 4b**; see[114] for an overview of unsupervised machine learning). In psychiatric conditions where behavioral manifestations are heterogeneous, these data-driven patterns might eventually help refine categories or tailor personalized interventions to individual-specific attentional profiles[115].

Unsupervised clustering based on features positions individuals in a shared feature space and defines the similarity among individuals with common distance measures (e.g., Euclidean distance, cosine distance). Same as during supervised classification, these features may include traditional AOI-based fixations, oculomotor features, or model-based phenotypes (e.g., visual saliency estimations, HMM-based gaze transition patterns). Each dimension of the space typically corresponds to one well-defined gaze feature, enabling high interpretability[51,116]. For example, one study performed hierarchical clustering based on seven attention measures (e.g., overall attention to a scene, to various parts of a person, and to distraction), revealing three distinct gaze subgroups within autistic toddlers (age 2): one subgroup featured high eye vs. mouth looking ratio, and one subgroup featured low eye vs. mouth looking ratio[116]. These three subtypes showed clinically relevant differences in skill acquisition rates during the subsequent year and predicted behavioral presentations at age 3. These findings support the use of unsupervised methods to parse heterogeneity of early syndrome expression, when standard diagnostic criteria may not yet be available. When many features are involved, dimensionality reduction such as PCA can help extract a more interpretable, low-dimensional latent space[17,117]. In a study that analyzed an initial total of 58 eye-tracking features, PCA revealed 3 dimensions that explained 59% of the total variance[117]. These three feature dimensions, representing the fixation duration (PC1), gaze step direction (PC2), and gaze step length (PC3), suggested two clusters of participants: a static viewer cluster with less frequent and longer fixations, and a dynamic viewer cluster with the opposite pattern[117]. As this example shows, combinations of computational methods condensed the rich feature set to yield compact and interpretable results.

The features can also be latent representations extracted from DNNs[118,119]. Although features extracted from these more complicated algorithms are generally less interpretable, they potentially yield more robust and reliable results. For instance, Elbattah et al. fed scanpath images from autistic and control participants to an autoencoder (a DNN used for dimensionality reduction) and generated low-dimensional vectors that revealed two clusters through subsequent K-means clustering[118]. In this study, other less sophisticated feature extraction methods, despite better interpretability, produced poorer separation of clusters. To enhance the interpretability of the latent representations generated with the autoencoder, the authors compared how the clusters were associated with several eye-tracking metrics, and found that elevated velocity was associated with a cluster mainly consisting of autistic individuals[118]. Another more recent study designed a novel DNN architecture that successfully predicted scanpaths for individual participants (with and without ASD) through the learning of observer-specific features[119]. Without providing the diagnostic labels to the model, the model-derived observer-specific features showed a natural separation of ASD and controls, suggesting the strong learning power of such models to discern unique gaze patterns in ASD[119].

Another unsupervised approach to consider is not based on features, but on the gaze trajectory itself. Feature-free gaze similarity typically relies on the inter-subject correlation (ISC) analysis, which measures the similarity of raw gaze data between individuals (see **Box 1**). ISC methods vary in focus[51,78,120–127]. Some methods emphasize the onset and direction of gaze shifts, such as the average Pearson correlation of the x and y time series[121,125,127]. Some methods emphasize the spatial proximities of the fixations, including correlations of fixation heatmaps[51,122], Euclidean distances between the average gaze positions (per time window)[123], and proportions of fixation overlaps[126]. Moreover, some scanpath comparison methods, such as the ScanMatch[128] and MultiMatch[129] algorithm, can quantify both spatial and temporal similarity in a single metric. Unlike feature-based methods that assign coordinates in a feature space, ISC reveals relative similarities across individuals. These in turn can be further analyzed with approaches that are model-free and require no dimensionality reduction at all, such as representational similarity analysis[130] (RSA, see **Box 1**). The relationships produced by ISC can also be visualized through various dimensionality reduction techniques (e.g., MDS[131], t-SNE[132], UMAP[133]) and may show clusters of participants that either confirm the hypothesized groups or suggest new subtypes.

One study used multi-dimensional scaling (MDS) to compare gaze to videos between typically developing children and children with ASD. Interestingly, instead of showing two clusters of participants (with vs. without ASD), the study found a similarity geometry that was characterized by a “core” (TD) vs. periphery” (ASD) structure, suggesting that TD children have higher gaze similarity to one another, while children with ASD deviate from TD in heterogeneous ways[134]. This type of approach is completely hypothesis-free and thus unbiased from any feature selection. Importantly, it helps address the challenge of interindividual heterogeneity: within a clinical group, individuals may differ from controls in opposite directions (some showing higher, others lower values), so the group average can appear similar to controls even when meaningful individual differences exist[135]. Such model-free similarity-based approaches have gained increasing attention in screening and diagnosis, since they can capture abnormalities by quantifying how an individual deviates from a normative sample in any direction[78,136,137].

Box 1. Application of Inter-Subject Correlation Analysis in Eye-Tracking

The idea of inter-subject correlation (ISC) was initially proposed in fMRI analysis[138], to provide a way to quantify the similarity in brain responses across individuals even for complex stimuli (movies) for which no simple model was available. In eye-tracking, ISC measures the spatial and/or temporal synchronization of eye movements between pairs of samples. Beyond serving as a similarity metric for unsupervised clustering (see main text), ISC offers specific insights valuable to psychiatry research.

Gaze heterogeneity across and within individuals

The most direct application of ISC characterizes how gaze to the same stimuli correlates between individuals. A simple extension quantifies how an individual’s gaze pattern deviates from the majority or mean of a group, reflecting that individual’s unique cognitive style and motivation. The results not only predict behavioral outcomes (e.g., individual learning success[124]) but often have clinical relevance. For example, several studies have found more idiosyncratic gaze patterns in children and adults with ASD[139]. The degree of this divergence correlates with ASD severity and other cognitive scores[51,123,126], as has been reported also for brain-derived ISCs[140,141].

In addition to inter-subject correlation, one can also measure intra-subject correlation, which is a meaningful index of the consistency of gaze patterns within an individual (over repeated tests). Similar to the between-individual heterogeneity discussed above, studies have also found associations between clinical symptoms and intra-subject gaze consistencies[51,123]. For example, reduced intra-subject consistency (less reliability across time) has been observed in children with ASD and is associated with lower inter-subject correlation compared to typically developing peers[123]; again, parallel results have also been reported in fMRI[142].

Time-resolved ISC analysis

Both the locations onto which we fixate as well as the time of such fixations are informative. One approach to better understand individual differences in spatio-temporal visual exploration patterns uses a 'moving time window' method to compute time-resolved ISC[122,127]. For example, a study that assessed ISC across monozygotic (MZ) and dizygotic (DZ) twins observed a time-varying ISC change: gaze locations across participants were more divergent within epochs of viewing complex natural scenes, and the level of divergence was higher in DZ than MZ pairs[122].

Time-resolved ISC is also helpful in analyzing video viewing data - to examine what type of visual features drives gaze convergence/divergence. This not only helps the design of more engaging videos for marketing and movie production, but also facilitates the design of clinical materials that better distinguish certain patient groups[78,120,123]. For instance, a study compared the ISC between several types of movies, and found that eye movements on Hollywood movie trailers were more coherent than those on movies of natural scenes[120]. Another study modeled time-resolved ISC between autistic and control individuals within a video as a function of the visual features[143]. Autistic participants were found to show greatest divergence from controls at moments when multiple faces co-existed in the scene.

Connecting multiple modalities using RSA

Representational similarity analysis provides a way to compare information across different data modalities by focusing not on the value of a measure, but on the similarity structure among patterns of responses[130,144–146]. It characterizes the geometry of how different stimuli, experimental conditions, or individuals relate to each other in representational space. This makes RSA a model-free approach that enables comparison across distinct data types, such as neural activity, eye-tracking, and behavioral responses. For example, gaze divergence across individuals (gaze ISC) has been found to predict neural divergence in the V1 and inferior temporal cortex (fMRI-based ISC) during movie watching[145]; and that differences in the way one describes a scene (description ISC) could also be explained by the differences in the way they look at the scene (gaze ISC)[146]. These results suggest that idiosyncrasy of gaze patterns drives distinct neural profiles and shapes unique subjective perceptions in response to the same visual stimuli.

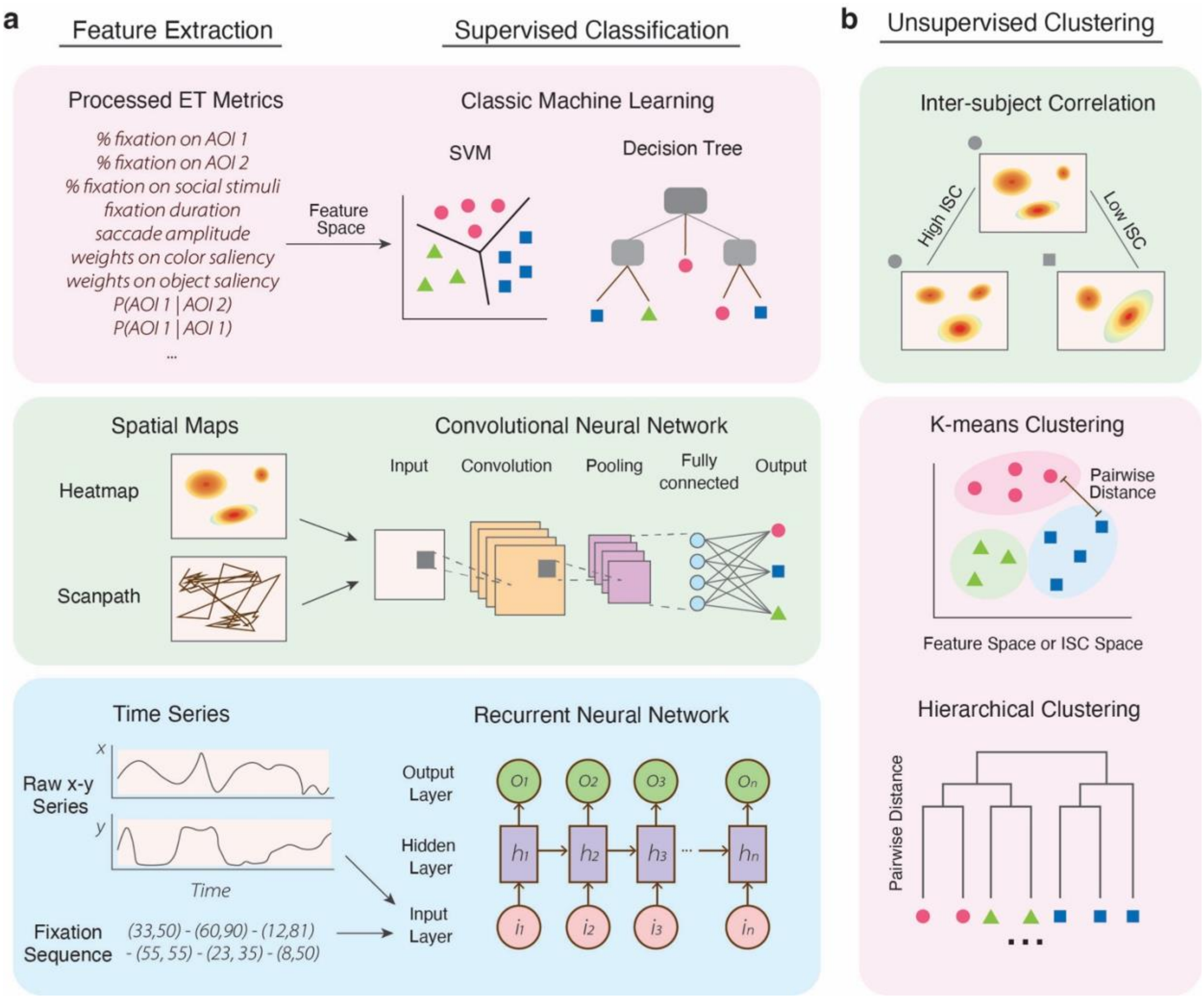


Figure 4. Machine learning in eye-tracking data analysis. (a) Supervised classification. Classic machine learning models can predict group labels with processed eye-tracking metrics. Convolutional neural networks are good at processing 2D images of gaze such as heatmap or scanpath and extracting lower-dimensional latent features for classification. Recurrent neural networks handle the temporal dependencies of time series data. Each dot represents a sample (one individual, or one individual's trial), and different shapes and colors represent different predefined classes. **(b)** Unsupervised clustering. When clustering is based on processed ET metrics, the (dis-)similarity can be defined as the distance (e.g., Euclidean distance) between samples in a feature space. When clustering based on the entire ET data, the similarity can be defined as the inter-subject correlation (ISC) and can be visualized in a low dimensional space. Clustering can be conducted given the similarities using different techniques, such as K-means clustering and hierarchical clustering. Each gray dot represents a sample that may come from different populations (different shapes), colors represent different data-driven clusters.

Advantages of using Computational Models in Eye-tracking Research

The four categories of methods we introduced (visual saliency model, data-driven AOI identification, supervised classification, unsupervised clustering) encompass a huge variation of theoretical bases, statistical assumptions, model classes, and ultimately - research goals. Nevertheless, they represent three key advantages that advance many types of psychiatry research: they inform psychological models of vision, provide richer and more sensitive measures for differentiating individuals, and make powerful clinical predictions.

First, the use of computational models strengthens our understanding of visual attention, and may suggest new hypotheses about attention mechanisms. Theory-based models, such as the visual saliency model, go beyond simple measures of fixation counts or dwell time by directly modeling multiple levels of features that guide gaze. ML models, such as CNNs, suggest image segments or visual features that contribute to distinguishing the fixation patterns between clinical groups. These approaches provide a more comprehensive characterization of spontaneous fixations in naturalistic settings and on complex stimuli, where multiple visual features co-exist and compete for attention (**Supplementary Information S2**). Unlike AOI-based analyses, many of these models are unbiased, without arbitrary AOI definitions, and readily applicable across clinical populations. Results from these models (e.g., saliency weight differences, image patches with high differentiability) can be mapped onto specific psychological processes, such as attention, control, or memory. As such, computational models can benefit the development of psychological theories of visual attention.

Second, more powerful models can help to characterize more fine-grained individual differences, some of which are not captured with traditional approaches. Several modern models are particularly helpful in extracting the temporal dynamics of gaze. Individuals who have the same dwell time on certain regions may have completely different gaze shift patterns (e.g., scan back and forth between two regions vs. look at each one for half the duration), which will be captured by EMHMM, RNN and ISC analyses. These richer metrics expand the dimensionality of analysis in biologically meaningful ways, offering saliency weights that quantify preferences to a wide array of features, and EMHMMs that generate fixation transition probabilities between AOIs. Such measures can serve as sensitive phenotypes for detecting clinically relevant individual variability that has been previously overlooked. Once such diagnostically meaningful constellations of visual features are discovered, they can inform the design of optimized stimuli that are tailored to differentiate clinical subgroups based on their unique attentional profiles. In the ideal case, one could iteratively use such an approach to generate eye-tracking stimuli that could be used in the clinic to screen for and to help diagnose psychiatric diseases. could be used in the clinic to screen for and to help diagnose psychiatric diseases.

Third, advanced machine learning algorithms enhance predictive and clinical utility. In addition to the success of supervised ML in classifying various clinical groups, some of the models show high interpretability and explain which eye movement features are most informative. Beyond classification, they have huge potential in tracking symptom progression and monitoring treatment efficacy. Unsupervised ML can even reveal transdiagnostic dimensions based on eye movement patterns, which benefits the design of individualized interventions within and across established diagnostic categories. Importantly, the predictive power of computational eye-tracking approaches can be further improved by integrating gaze with other modalities[79,94], such as eye blinks, pupillometry, behavioral assessments, neuroimaging, physiological signals, or genotypes[79,94,147,148]. For instance, a mobile app developed for early detection of autism achieved strong accuracy with gaze features alone, and even higher accuracy when combined with facial expressions, head movements, and user interaction metrics – yet gaze remained one of the top predictive features[79]. As multimodal fusion advances, including through large language models[147,148] , eye-tracking will contribute significantly to precision psychiatry.

Future Directions and Open Questions

Establishing standardized procedures

A perennial challenge in eye-tracking research is the relatively low consistency of findings across studies, even when testing the same construct on the same clinical population[149,150]. Part of the inconsistency stems from the lack of standardized procedures in eye-tracking experiments, spanning equipment (see discussions of advanced eye-tracking technologies in **Supplementary Information S3**), testing environment, quality control, data processing and documentation. Standardization becomes even more challenging in real-world protocols, where variability in lighting, stimulus features, and motion can further hinder comparability across studies (despite its benefit to ecological validity, see **Supplementary Information S2**). To mitigate this issue, recent guidelines argue for a set of parameters to report during data collection and data processing of eye-tracking research[151].

Although modern eye trackers offer high spatial and temporal resolution[152], the data remain susceptible to noise. Excessive blinks, head movement, and off-screen gazes can cause data loss or inaccurate gaze estimation[153]. Because such behaviors may be associated with psychiatric symptoms, the resulting noise can confound analyses. Computational models might misinterpret noise as meaningful patterns and show spurious correlations with variables of interest. Therefore, careful quality checks and clear justification of inclusion and exclusion criteria are essential. Ideally, results are computed across a range of data quality, ensuring that no spurious findings emerge as data quality changes. When deriving individual phenotypes for group-level comparisons, one should assess model fits for each participant. The quality of the model fit can itself be informative about the participant's condition, yet poorly fitted models can introduce substantial noise if their parameters are used in further analyses. For novel mathematical models that have yet to be validated, additional steps - such as model simulations, parameter recovery, model comparisons, and posterior predictive checks - are essential (see a practical guide of best practices in model validation[154]). These best practices build a solid foundation for integrating eye-tracking data across sites, studies, and modalities.

Building larger and richer datasets

With standardized methods in place, the next key priority is to assemble larger and more representative eye-tracking datasets. Many existing studies in psychiatric research are limited by small sample sizes and lack normative data stratified by demographics, impeding statistical power, replicability, and clinical translation. When eye-tracking metrics are proposed as diagnostic biomarkers, the lack of demographic-specific normative references hinders their translation to clinical use. Establishing large normative datasets that capture the typical range of gaze characteristics across age, sex, and other demographic factors is therefore essential. Early efforts in this direction, such as mapping developmental trajectories of attention biases across sexes, have illustrated the potential of such resources in providing reliable reference points[155,156].

The need for larger datasets is further amplified by the use of computational models and machine learning. These approaches involve numerous parameters and require extensive data, both in terms of longer recordings for individualized modeling and larger samples for classification or clustering. Machine learning models trained on small and homogeneous datasets are prone to overfitting and may fail to generalize to new samples. Multi-site collaborations and open data sharing are making large-scale data curation increasingly feasible. To ensure generalizability, future studies should train models on combined datasets collected from diverse populations and sites, while maintaining open sharing practices that allow cumulative sample growth and cross-study comparison.

Toward better clinical translation

With standardized methods and large, demographically diverse datasets in place, the next step is to translate computational eye-tracking findings into clinically meaningful applications. Although computational phenotypes and machine learning models show promise for classifying psychiatric groups based on gaze patterns, their diagnostic and predictive validity remains uncertain.

For case-control group comparisons, results are fundamentally shaped by how groups are defined to begin with. For example, differences observed between healthy controls and patients who have only one psychiatric diagnosis without comorbidity may not generalize when comparing with patients who have other comorbid symptoms. Similarly, predictive algorithm trained on one demographic group may not generalize to others. Much of the existing work focuses on binary classifications against healthy controls. This leaves unclear the discriminant validity of findings when compared to other clinical groups and overlooks the complexity of psychiatric diagnosis, where overlapping symptoms are common across conditions. Future efforts should expand beyond control-versus-patient contrasts to include multiple patient groups with overlapping functional challenges (e.g., autism and social anxiety in social attention[5,157,158], depression and anxiety in attention to emotional stimuli[150,159]). In fact, lack of discriminant validity for related psychiatric conditions might suggest a transdiagnostic nature of the psychological mechanism underlying a phenotype: disorders may share common processes, and interventions aimed at transdiagnostic dimensions may be effective for multiple diagnoses[160]. Understanding whether a phenotype is shared across diagnoses or unique to a particular condition can validate or revise the current diagnosis classification, informing decisions about screening, diagnosis, and treatment that may be targeted at one or more clinical groups.

A second major obstacle for clinical translation is the lack of explainability of advanced computational models. Deep neural network models, in particular, are often referred to as ‘black boxes’, because how they work and what information they use when making predictions are undisclosed. Although their performance can surpass that of human experts, the opacity of judgements poses ethical challenges: if clinicians cannot understand the decision-making, they will not be able to communicate this to patients, thus affecting patients’ ability to engage in informed consent[161]. Furthermore, some high-performing models may accidentally achieve high accuracy by using confounding variables, which can eventually lead to misleading or even harmful outcomes[162]. The pursuit of explainable AI therefore remains critical, and ongoing research has made efforts to address the concern by identifying features and layers of neural network that contribute to the prediction, offering a path toward safer and more trustworthy applications.

Conclusion

Eye movement patterns are quantifiable, difficult to consciously control, and reflect neural processing of attention and perception. Many of the clinically relevant conclusions that have been drawn from neuroimaging studies are mirrored in eye-tracking studies – for a fraction of the cost and substantially greater ease of application. Recent advances in machine learning and artificial intelligence, along with broader computational methods, have offered new ways of collecting and analyzing eye-tracking data to realize its full potential in psychiatry research.

Visual saliency models that predict gaze patterns based on low-level visual properties have been expanded to incorporate multiple levels of features, such as objects and semantic content. This has led to improved accuracy in predicting human gaze and greater sensitivity to individual differences in attention patterns. Data-driven, multivariate characterizations of gaze complement hypothesis-driven analyses by clustering parts of the stimuli that can be meaningfully grouped or by identifying distinct attention “styles” across individuals. Both supervised and unsupervised machine learning methods can utilize multiple eye-tracking metrics to reveal gaze patterns associated with a particular psychiatric condition, or can identify subtypes within or across clinical conditions.

Finally, these analytic approaches offer promising new directions for future research– such as combining these methods with scalable, camera-based eye-tracking through smartphones or webcams; applying novel experimental tools to probe gaze patterns during real-life interactions; and leveraging large-scale databases of clinical conditions with substantial sample sizes. Thus, computational approaches to eye-tracking research holds significant potential to reveal individual differences and to advance our understanding of cognition in both the general population and among psychiatric conditions.

Acknowledgement

This study was funded by the Simons Foundation Autism Research Initiative grant (grant # 990500).

Author’s Contributions

Q.W. and R.A. conceived the study. Q.W. and N.K. conducted the literature review and drafted the initial manuscript. R.A. provided substantial feedback and edits to the original draft. All authors reviewed and revised the manuscript and approved the final version.

Competing Interests

The authors declare no competing interests.

Reference

1. Andrews, T. J. & Coppola, D. M. Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research* **39**, 2947–2953 (1999).
2. Henderson, J. M. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences* **7**, 498–504 (2003).
3. Otero-Millan, J., Macknik, S. L., Langston, R. E. & Martinez-Conde, S. An oculomotor continuum from exploration to fixation. *Proceedings of the National Academy of Sciences* **110**, 6175–6180 (2013).
4. Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I. & Sherman, A. M. Visual search for arbitrary objects in real scenes. *Atten Percept Psychophys* **73**, 1650–1671 (2011).
5. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344 (2017).
6. Hoffman, J. E. Visual Attention and Eye Movements. in *Attention* (Psychology Press, 1998).
7. Rahal, R.-M. & Fiedler, S. Understanding cognitive and affective mechanisms in social psychology through eye-tracking. *Journal of Experimental Social Psychology* **85**, 103842 (2019).

8. Orquin, J. L. & Mueller Loose, S. Attention and choice: A review on eye movements in decision making. *Acta Psychologica* **144**, 190–206 (2013).

9. Hannula, D. E. *et al.* Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Front Hum Neurosci* **4**, 166 (2010).

10. Coiner, B. *et al.* Functional neuroanatomy of the human eye movement network: a review and atlas. *Brain Struct Funct* **224**, 2603–2617 (2019).

11. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu Rev Neurosci* **18**, 193–222 (1995).

12. Kim, N. Y. & Kastner, S. A biased competition theory for the developmental cognitive neuroscience of visuo-spatial attention. *Current Opinion in Psychology* **29**, 219–228 (2019).

13. Greenlee, M. W. & Kimmig, H. Visual Perception and Eye Movements. in *Eye Movement Research: An Introduction to its Scientific Foundations and Applications* (eds Klein, C. & Ettinger, U.) 165–196 (Springer International Publishing, Cham, 2019). doi:10.1007/978-3-030-20085-5_5.

14. Itti, L. & Koch, C. Computational modelling of visual attention. *Nat Rev Neurosci* **2**, 194–203 (2001).

15. Wang, S. *et al.* Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron* **88**, 604–616 (2015).

16. Haskins, A. J. *et al.* Reduced social attention in autism is magnified by perceptual load in naturalistic environments. *Autism Research* **15**, 2310–2323 (2022).

17. Nayar, K., Shic, F., Winston, M. & Losh, M. A constellation of eye-tracking measures reveals social attention differences in ASD and the broad autism phenotype. *Molecular Autism* **13**, 18 (2022).

18. Yiend, J. & Mathews, A. Anxiety and attention to threatening pictures. *The Quarterly Journal of Experimental Psychology Section A* **54**, 665–681 (2001).

19. MacLeod, C. & Mathews, A. Anxiety and the Allocation of Attention to Threat. *The Quarterly Journal of Experimental Psychology Section A* **40**, 653–670 (1988).

20. Türkan, B. N., Amado, S., Ercan, E. S. & Perçinel, I. Comparison of change detection performance and visual search patterns among children with/without ADHD: Evidence from eye movements. *Res Dev Disabil* **49–50**, 205–215 (2016).

21. Caldani, S. *et al.* The Effect of Dual Task on Attentional Performance in Children With ADHD. *Front. Integr. Neurosci.* **12**, (2019).

22. Bischof, W. F., Anderson, N. C. & Kingstone, A. Temporal Methods for Eye Movement Analysis. in *Eye Movement Research* 407–448 (Springer, Cham, 2019). doi:10.1007/978-3-030-20085-5_10.

23. Oberwelland, E. *et al.* Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *NeuroImage* **130**, 248–260 (2016).

24. Griffin, J. W. *et al.* Spatiotemporal Eye Movement Dynamics Reveal Altered Face Prioritization in Early Visual Processing Among Autistic Children. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* https://doi.org/10.1016/j.bpsc.2024.08.017 (2024) doi:10.1016/j.bpsc.2024.08.017.

25. Veale, R., Hafed, Z. M. & Yoshida, M. How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling. *Phil. Trans. R. Soc. B* **372**, 20160113 (2017).

26. Koch, C. & Ullman, S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. in *Matters of Intelligence: Conceptual Structures in Cognitive Neuroscience* (ed. Vaina, L. M.) 115–141 (Springer Netherlands, Dordrecht, 1987). doi:10.1007/978-94-009-3833-5_5.

27. Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cognitive Psychology* **12**, 97–136 (1980).

28. Itti, L., Koch, C. & Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 1254–1259 (1998).

29. Judd, T., Ehinger, K., Durand, F. & Torralba, A. Learning to predict where humans look. in *2009 IEEE 12th International Conference on Computer Vision* 2106–2113 (2009). doi:10.1109/ICCV.2009.5459462.

30. Harel, J., Koch, C. & Perona, P. Graph-Based Visual Saliency. in *Advances in Neural Information Processing Systems* vol. 19 (MIT Press, 2006).

31. Hou, X. & Zhang, L. Saliency Detection: A Spectral Residual Approach. in *2007 IEEE Conference on Computer Vision and Pattern Recognition* 1–8 (2007). doi:10.1109/CVPR.2007.383267.

32. Zhang, L., Tong, M. H., Marks, T. K., Shan, H. & Cottrell, G. W. SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision* **8**, 32 (2008).

33. Linardos, A., Kümmerer, M., Press, O. & Bethge, M. DeepGaze IIE: Calibrated Prediction in and Out-of-Domain for State-of-the-Art Saliency Modeling. in 12919–12928 (2021).

34. Huang, X., Shen, C., Boix, X. & Zhao, Q. SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks. in 262–270 (2015).

35. Kruthiventi, S. S. S., Ayush, K. & Babu, R. V. DeepFix: A Fully Convolutional Neural Network for Predicting Human Eye Fixations. *IEEE Trans Image Process* **26**, 4446–4456 (2017).

36. Kümmerer, M. *et al.* MIT/Tübingen Saliency Benchmark.

37. Jain, S. *et al.* ViNet: Pushing the limits of Visual Modality for Audio-Visual Saliency Prediction. Preprint at https://doi.org/10.48550/arXiv.2012.06170 (2021).

38. Tavakoli, H. R., Borji, A., Rahtu, E. & Kannala, J. DAVE: A Deep Audio-Visual Embedding for Dynamic Saliency Prediction. Preprint at https://doi.org/10.48550/arXiv.1905.10693 (2020).

39. Kümmerer, M., Bethge, M. & Wallis, T. S. A. DeepGaze III: Modeling free-viewing human scanpaths with deep learning. *Journal of Vision* **22**, 7 (2022).

40. Roth, N., Rolfs, M., Hellwich, O. & Obermayer, K. Objects guide human gaze behavior in dynamic real-world scenes. *PLOS Computational Biology* **19**, e1011512 (2023).

41. Mengers, V., Roth, N., Brock, O., Obermayer, K. & Rolfs, M. A robotics-inspired scanpath model reveals the importance of uncertainty and semantic object cues for gaze guidance in dynamic scenes. *Journal of Vision* **25**, 6 (2025).

42. Zanca, D., Melacci, S. & Gori, M. Gravitational Laws of Focus of Attention. *IEEE Trans Pattern Anal Mach Intell* **42**, 2983–2995 (2020).

43. Kümmerer, M. & Bethge, M. State-of-the-Art in Human Scanpath Prediction. Preprint at https://doi.org/10.48550/arXiv.2102.12239 (2021).

44. Judd, T., Durand, F. & Torralba, A. A Benchmark of Computational Models of Saliency to Predict Human Fixations. https://dspace.mit.edu/handle/1721.1/68590 (2012).

45. Borji, A. & Itti, L. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research. Preprint at https://doi.org/10.48550/arXiv.1505.03581 (2015).

46. Strauch, C. *et al.* Saliency models perform best for women's and young adults' fixations. *Commun Psychol* **1**, 1–10 (2023).

47. Adámek, P. *et al.* The Gaze of Schizophrenia Patients Captured by Bottom-up Saliency. *Schizophr* **10**, 1–13 (2024).

48. Bast, N. *et al.* Sensory salience processing moderates attenuated gazes on faces in autism spectrum disorder: a case–control study. *Molecular Autism* **14**, 5 (2023).

49. Dziemian, S. *et al.* Saliency Models Reveal Reduced Top-Down Attention in Attention-Deficit/Hyperactivity Disorder: A Naturalistic Eye-Tracking Study. *JAACAP Open* S2949732924000280 (2024) doi:10.1016/j.jaacop.2024.03.001.

50. de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S. & Gegenfurtner, K. R. Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences* **116**, 11687–11692 (2019).

51. Keles, U. *et al.* Atypical gaze patterns in autistic adults are heterogeneous across but reliable within individuals. *Molecular Autism* **13**, 39 (2022).

52. Król, M. E. & Król, M. A novel machine learning analysis of eye-tracking data reveals suboptimal visual information extraction from facial stimuli in individuals with autism. *Neuropsychologia* **129**, 397–406 (2019).

53. Zhang, W., Li, X., Liu, X., Lu, S. & Tang, H. Facing challenges: A survey of object tracking. *Digital Signal Processing* **161**, 105082 (2025).

54. Vogg, R. *et al.* Computer vision for primate behavior analysis in the wild. *Nat Methods* **22**, 1154–1166 (2025).

55. Uneza, Gupta, D. & Saini, S. Facial Expression Analysis: Unveiling the Emotions Through Computer Vision. in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* 1–5 (2024). doi:10.1109/ICRITO61523.2024.10522418.

56. Haskins, A. J., Mentch, J., Botch, T. L. & Robertson, C. E. Active vision in immersive, 360° real-world environments. *Sci Rep* **10**, 14304 (2020).

57. Haskins, A. J., Mentch, J., Van Wicklin, C., Choi, Y. B. & Robertson, C. E. Brief Report: Differences in Naturalistic Attention to Real-World Scenes in Adolescents with 16p.11.2 Deletion. *J Autism Dev Disord* **54**, 1078–1087 (2024).

58. Dalrymple, K. A., Jiang, M., Zhao, Q. & Elison, J. T. Machine learning accurately classifies age of toddlers based on eye tracking. *Sci Rep* **9**, 6255 (2019).

59. Henderson, J. M. & Hayes, T. R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat Hum Behav* **1**, 743–747 (2017).

60. Broda, M. D. & de Haas, B. Individual differences in human gaze behavior generalize from faces to objects. *Proceedings of the National Academy of Sciences* **121**, e2322149121 (2024).

61. Falck-Ytter, T. The breakdown of social looking. *Neuroscience & Biobehavioral Reviews* **161**, 105689 (2024).

62. Broda, M. D. & De Haas, B. Individual differences in looking at persons in scenes. *Journal of Vision* **22**, 9 (2022).

63. Wegner-Clemens, K., Rennig, J., Magnotti, J. F. & Beauchamp, M. S. Using principal component analysis to characterize eye movement fixation patterns during face viewing. *Journal of Vision* **19**, 2 (2019).

64. Liu, W., Li, M. & Yi, L. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research* **9**, 888–898 (2016).

65. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nat Hum Behav* **1**, 0021 (2017).

66. Masulli, P. *et al.* Data-driven analysis of gaze patterns in face perception: Methodological and clinical contributions. *Cortex* **147**, 9–23 (2022).

67. Klötzl, D. *et al.* NMF-Based Analysis of Mobile Eye-Tracking Data. in 1–9 (Association for Computing Machinery, New York, NY, USA, 2024). doi:10.1145/3649902.3653518.

68. Hsiao, J. H., Lan, H., Zheng, Y. & Chan, A. B. Eye movement analysis with hidden Markov models (EMHMM) with co-clustering. *Behav Res* **53**, 2473–2486 (2021).

69. Hsiao, J. H., An, J., Hui, V. K. S., Zheng, Y. & Chan, A. B. Understanding the role of eye movement consistency in face recognition and autism through integrating deep neural networks and hidden Markov models. *npj Sci. Learn.* **7**, 1–13 (2022).

70. Hsiao, J. H. Understanding Human Cognition Through Computational Modeling. *Topics in Cognitive Science* **16**, 349–376 (2024).

71. Chuk, T., Chan, A. B. & Hsiao, J. H. Understanding eye movements in face recognition using hidden Markov models. *Journal of Vision* **14**, 8 (2014).

72. Coviello, E., Chan, A. B. & Lanckriet, G. R. G. Clustering Hidden Markov Models with Variational HEM. *Journal of Machine Learning Research* **15**, 697–747 (2014).

73. Chan, C. Y. H., Chan, A. B., Lee, T. M. C. & Hsiao, J. H. Eye-movement patterns in face recognition are associated with cognitive decline in older adults. *Psychon Bull Rev* **25**, 2200–2207 (2018).

74. Karvelis, P., Paulus, M. P. & Diaconescu, A. O. Individual differences in computational psychiatry: A review of current challenges. *Neuroscience & Biobehavioral Reviews* **148**, 105137 (2023).

75. Hauser, T. U., Skvortsova, V., De Choudhury, M. & Koutsouleris, N. The promise of a model-based psychiatry: building computational models of mental ill health. *The Lancet Digital Health* **4**, e816–e828 (2022).

76. Washington, P. *et al.* Data-Driven Diagnostics and the Potential of Mobile Artificial Intelligence for Digital Therapeutic Phenotyping in Computational Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **5**, 759–769 (2020).

77. Jones, W. *et al.* Eye-Tracking–Based Measurement of Social Visual Engagement Compared With Expert Clinical Diagnosis of Autism. *JAMA* **330**, 854–865 (2023).

78. Jones, W. *et al.* Development and Replication of Objective Measurements of Social Visual Engagement to Aid in Early Diagnosis and Assessment of Autism. *JAMA Network Open* **6**, e2330145 (2023).

79. Perochon, S. *et al.* Early detection of autism using digital behavioral phenotyping. *Nat Med* **29**, 2489–2497 (2023).

80. Nazari, S. *et al.* Large-scale examination of early-age sex differences in neurotypical toddlers and those with autism spectrum disorder or other developmental conditions. *Nat Hum Behav* **9**, 1697–1709 (2025).

81. O’Driscoll, G. A. & Callahan, B. L. Smooth pursuit in schizophrenia: A meta-analytic review of research since 1993. *Brain and Cognition* **68**, 359–370 (2008).

82. Athanasopoulos, F., Saprikis, O.-V., Margeli, M., Klein, C. & Smyrnis, N. Towards Clinically Relevant Oculomotor Biomarkers in Early Schizophrenia. *Front. Behav. Neurosci.* **15**, (2021).

83. Shishido, E. *et al.* Application of eye trackers for understanding mental disorders: Cases for schizophrenia and autism spectrum disorder. *Neuropsychopharmacology Reports* **39**, 72–77 (2019).

84. Pargent, F., Schoedel, R. & Stachl, C. Best Practices in Supervised Machine Learning: A Tutorial for Psychologists. *Advances in Methods and Practices in Psychological Science* **6**, 25152459231162559 (2023).

85. Alcañiz, M. *et al.* Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: A proof of concept for diagnosis. *Autism Research* **15**, 131–145 (2022).

86. Benson, P. J. *et al.* Simple Viewing Tests Can Detect Eye Movement Abnormalities That Distinguish Schizophrenia Cases from Controls with Exceptional Accuracy. *Biological Psychiatry* **72**, 716–724 (2012).

87. Lee, D. Y. *et al.* Use of eye tracking to improve the identification of attention-deficit/hyperactivity disorder in children. *Sci Rep* **13**, 14469 (2023).

88. Li, J. *et al.* Classifying ASD children with LSTM based on raw videos. *Neurocomputing* **390**, 226–238 (2020).

89. Zheng, Z. *et al.* Diagnosing and tracking depression based on eye movement in response to virtual reality. *Front Psychiatry* **15**, 1280935 (2024).

90. Kim, M. *et al.* Development of an eye-tracking system based on a deep learning model to assess executive function in patients with mental illnesses. *Sci Rep* **14**, 18186 (2024).

91. Kacur, J., Polec, J., Smolejova, E. & Heretik, A. An Analysis of Eye-Tracking Features and Modelling Methods for Free-Viewed Standard Stimulus: Application for Schizophrenia Detection. *IEEE Journal of Biomedical and Health Informatics* **24**, 3055–3065 (2020).

92. Wei, Q., Cao, H., Shi, Y., Xu, X. & Li, T. Machine learning based on eye-tracking data to identify Autism Spectrum Disorder: A systematic review and meta-analysis. *Journal of Biomedical Informatics* **137**, 104254 (2023).

93. Mendez-Encinas, D., Sujar, A., Bayona, S. & Delgado-Gomez, D. Attention and impulsivity assessment using virtual reality games. *Sci Rep* **13**, 13689 (2023).

94. Wiebe, A. *et al.* Virtual reality-assisted prediction of adult ADHD based on eye tracking, EEG, actigraphy and behavioral indices: a machine learning analysis of independent training and test samples. *Transl Psychiatry* **14**, 508 (2024).

95. Revers, M. C. *et al.* Classification of Autism Spectrum Disorder Severity Using Eye Tracking Data Based on Visual Attention Model. in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)* 142–147 (2021). doi:10.1109/CBMS52027.2021.00062.

96. Tseng, P.-H. *et al.* High-throughput classification of clinical populations from natural viewing eye movements. *J Neurol* **260**, 275–284 (2013).

97. Coutrot, A., Hsiao, J. H. & Chan, A. B. Scanpath modeling and classification with hidden Markov models. *Behav Res* **50**, 362–379 (2018).

98. Fisher, R. A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* **7**, 179–188 (1936).

99. Shannon, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal* **27**, 379–423 (1948).

100. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* **46**, 389–422 (2002).

101. Peng, H., Long, F. & Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, 1226–1238 (2005).

102. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 4768–4777 (Curran Associates Inc., Red Hook, NY, USA, 2017).

103. Ribeiro, M. T., Singh, S. & Guestrin, C. 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135–1144 (Association for Computing Machinery, New York, NY, USA, 2016). doi:10.1145/2939672.2939778.

104. Kanhirakadavath, M. R. & Chandran, M. S. M. Investigation of Eye-Tracking Scan Path as a Biomarker for Autism Screening Using Machine Learning Algorithms. *Diagnostics (Basel)* **12**, 518 (2022).

105. Jaradat, A. S., Wedyan, M., Alomari, S. & Barhoush, M. M. Using Machine Learning to Diagnose Autism Based on Eye Tracking Technology. *Diagnostics (Basel)* **15**, 66 (2024).

106. Deng, S. *et al.* Detection of ADHD Based on Eye Movements During Natural Viewing. in *Machine Learning and Knowledge Discovery in Databases* (eds Amini, M.-R. et al.) vol. 13718 403–418 (Springer Nature Switzerland, Cham, 2023).

107. Song, Y. *et al.* EMS: A Large-Scale Eye Movement Dataset, Benchmark, and New Model for Schizophrenia Recognition. *IEEE Transactions on Neural Networks and Learning Systems* 1–12 (2024) doi:10.1109/TNNLS.2024.3441928.

108. Jiang, M. & Zhao, Q. Learning Visual Attention to Identify People with Autism Spectrum Disorder. in *2017 IEEE International Conference on Computer Vision (ICCV)* 3287–3296 (IEEE, Venice, 2017). doi:10.1109/ICCV.2017.354.

109. Ahmed, Z. A. T. *et al.* Applying Eye Tracking with Deep Learning Techniques for Early-Stage Detection of Autism Spectrum Disorders. *Data* **8**, 168 (2023).

110. Kacur, J., Polec, J., Smolejova, E. & Heretik, A. An Analysis of Eye-Tracking Features and Modelling Methods for Free-Viewed Standard Stimulus: Application for Schizophrenia Detection. *IEEE J Biomed Health Inform* **24**, 3055–3065 (2020).

111. Tao, Y. & Shyu, M.-L. SP-ASDNet: CNN-LSTM Based ASD Classification Model using Observer ScanPaths. *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* 641–646 (2019) doi:10.1109/ICMEW.2019.00124.

112. Cheekaty, S. & Muneeswari, G. Enhanced multilevel autism classification for children using eye-tracking and hybrid CNN-RNN deep learning models. *Neural Comput & Applic* https://doi.org/10.1007/s00521-024-10633-0 (2024) doi:10.1007/s00521-024-10633-0.

113. Elmadjian, C., Gonzales, C., Costa, R. L. da & Morimoto, C. H. Online eye-movement classification with temporal convolutional networks. *Behav Res* **55**, 3602–3620 (2023).

114. Gao, C. X. *et al.* An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research* **327**, 115265 (2023).

115. Wang, Q. *et al.* Interactive eye tracking for gaze strategy modification. in *Proceedings of the 14th International Conference on Interaction Design and Children* 247–250 (Association for Computing Machinery, New York, NY, USA, 2015). doi:10.1145/2771839.2771888.

116. Campbell, D. J., Shic, F., Macari, S. & Chawarska, K. Gaze Response to Dyadic Bids at 2 Years Related to Outcomes at 3 Years in Autism Spectrum Disorders: A Subtyping Analysis. *J Autism Dev Disord* **44**, 431–442 (2014).

117. Zangrossi, A., Cona, G., Celli, M., Zorzi, M. & Corbetta, M. Visual exploration dynamics are low-dimensional and driven by intrinsic factors. *Commun Biol* **4**, 1100 (2021).

118. Elbattah, M., Carette, R., Dequen, G., Guérin, J.-L. & Cilia, F. Learning Clusters in Autism Spectrum Disorder: Image-Based Clustering of Eye-Tracking Scanpaths with Deep Autoencoder. in *2019 41st Annual International Conference of the IEEE Engineering in*

Medicine and Biology Society (EMBC) 1417–1420 (2019). doi:10.1109/EMBC.2019.8856904.

119. Chen, X., Jiang, M. & Zhao, Q. Beyond Average: Individualized Visual Scanpath Prediction. in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 25420–25431 (IEEE, Seattle, WA, USA, 2024). doi:10.1109/CVPR52733.2024.02402.
120. Dorr, M., Martinetz, T., Gegenfurtner, K. R. & Barth, E. Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision* **10**, 28–28 (2010).
121. Shepherd, S. V., Steckenfinger, S. A., Hasson, U. & Ghazanfar, A. A. Human-Monkey Gaze Correlations Reveal Convergent and Divergent Patterns of Movie Viewing. *Current Biology* **20**, 649–656 (2010).
122. Kennedy, D. P. *et al.* Genetic Influence on Eye Movements to Complex Scenes at Short Timescales. *Current Biology* **27**, 3554-3560.e3 (2017).
123. Avni, I. *et al.* Children with autism observe social interactions in an idiosyncratic manner. *Autism Research* **13**, 935–946 (2020).
124. Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R. & Parra, L. C. Synchronized eye movements predict test scores in online video education. *Proceedings of the National Academy of Sciences* **118**, e2016980118 (2021).
125. Gu, C., Peng, Y., Nastase, S. A., Mayer, R. E. & Li, P. Onscreen presence of instructors in video lectures affects learners’ neural synchrony and visual attention during multimedia learning. *Proceedings of the National Academy of Sciences* **121**, e2309054121 (2024).
126. Hou, W., Cheng, R., Zhao, Z., Liao, H. & Li, J. Atypical and variable attention patterns reveal reduced contextual priors in children with autism spectrum disorder. *Autism Research* **17**, 1572–1585 (2024).
127. Hedger, N. & Chakrabarti, B. Autistic differences in the temporal dynamics of social attention. *Autism* **25**, 1615–1626 (2021).

128. Cristino, F., Mathôt, S., Theeuwes, J. & Gilchrist, I. D. ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods* **42**, 692–700 (2010).

129. Dewhurst, R. *et al.* It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behav Res* **44**, 1079–1100 (2012).

130. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, (2008).

131. Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika* **17**, 401–419 (1952).

132. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

133. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://doi.org/10.48550/arXiv.1802.03426 (2020).

134. Nakano, T. *et al.* Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proc. R. Soc. B.* **277**, 2935–2943 (2010).

135. Segal, A. *et al.* Embracing variability in the search for biological mechanisms of psychiatric illness. *Trends in Cognitive Sciences* **29**, 85–99 (2025).

136. Wolfers, T. *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry* **75**, 1146–1155 (2018).

137. Ge, R. *et al.* Normative modelling of brain morphometry across the lifespan with CentileBrain: algorithm benchmarking and model optimisation. *The Lancet Digital Health* **6**, e211–e221 (2024).

138. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634–1640 (2004).

139. Franchak, J. M., Heeger, D. J., Hasson, U. & Adolph, K. E. Free Viewing Gaze Behavior in Infants and Adults. *Infancy* **21**, 262–287 (2016).

140. Byrge, L., Dubois, J., Tyszka, J. M., Adolphs, R. & Kennedy, D. P. Idiosyncratic Brain Activation Patterns Are Associated with Poor Social Comprehension in Autism. *J. Neurosci.* **35**, 5837–5850 (2015).

141. Hasson, U. *et al.* Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res* **2**, 220–231 (2009).

142. Dinstein, I. *et al.* Unreliable evoked responses in autism. *Neuron* **75**, 981–991 (2012).

143. Wu, Q., Kim, N. Y., Turner, J. M., Paul, L. K. & Adolphs, R. Modeling Eye Gaze to Videos Using Dynamic Trajectory Variability Analysis. *INSAR 2023* https://www.biologicalpsychiatryjournal.com/article/S0006-3223(23)00470-5/abstract (2023).

144. Kiat, J. E. *et al.* Linking patterns of infant eye movements to a neural network model of the ventral stream using representational similarity analysis. *Developmental Science* **25**, e13155 (2022).

145. Borovska, P. & de Haas, B. Individual gaze shapes diverging neural representations. *Proceedings of the National Academy of Sciences* **121**, e2405602121 (2024).

146. Kollenda, D., Reher, A.-S. & de Haas, B. Individual gaze predicts individual scene descriptions. *Sci Rep* **15**, 9443 (2025).

147. Ma, C. C. *et al.* Multimodal Fusion with LLMs for Engagement Prediction in Natural Conversation. Preprint at https://doi.org/10.48550/ARXIV.2409.09135 (2024).

148. Wu, M. *et al.* Hypergraph Multi-modal Large Language Model: Exploiting EEG and Eye-tracking Modalities to Evaluate Heterogeneous Responses for Video Understanding. in 7316–7325 (ACM, Melbourne VIC Australia, 2024). doi:10.1145/3664647.3680810.

149. Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B. & Lagopoulos, J. A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience* **9**, 610–632 (2014).

150. Suslow, T., Hußlack, A., Kersting, A. & Bodenschatz, C. M. Attentional biases to emotional information in clinical depression: A systematic and meta-analytic review of eye tracking findings. *Journal of Affective Disorders* **274**, 632–642 (2020).

151. Dunn, M. J. *et al.* Minimal reporting guideline for research involving eye tracking (2023 edition). *Behav Res Methods* **56**, 4351–4357 (2024).

152. Funke, G. *et al.* Which Eye Tracker Is Right for Your Research? Performance Evaluation of Several Cost Variant Eye Trackers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **60**, 1240–1244 (2016).

153. Wass, S. V., Forssman, L. & Leppänen, J. Robustness and Precision: How Data Quality May Influence Key Dependent Variables in Infant Eye-Tracker Analyses. *Infancy* **19**, 427–460 (2014).

154. Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**, e49547 (2019).

155. Linka, M., Karimpur, H. & de Haas, B. Protracted development of gaze behaviour. *Nat Hum Behav* 1–11 (2025) doi:10.1038/s41562-025-02191-9.

156. Strauch, C., Hoogerbrugge, A. J. & Ten Brink, A. F. Gaze data of 4243 participants shows link between leftward and superior attention biases and age. *Exp Brain Res* **242**, 1327–1337 (2024).

157. Kleberg, J. L. *et al.* Autistic Traits and Symptoms of Social Anxiety are Differentially Related to Attention to Others' Eyes in Social Anxiety Disorder. *J Autism Dev Disord* **47**, 3814–3821 (2017).

158. Mansell, W., Clark ,David M., Ehlers ,Anke & and Chen, Y.-P. Social Anxiety and Attention away from Emotional Faces. *Cognition and Emotion* **13**, 673–690 (1999).

159. Armstrong, T. & Olatunji, B. O. Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review* **32**, 704–723 (2012).

160. Fusar-Poli, P. *et al.* Transdiagnostic psychiatry: a systematic review. *World Psychiatry* **18**, 192–207 (2019).

161. Xu, H. & Shuttleworth, K. M. J. Medical artificial intelligence and the black box problem: a view based on the ethical principle of “do no harm”. *Intelligent Medicine* **4**, 52–57 (2024).

162. Reddy, S. Explainability and artificial intelligence in medicine. *The Lancet Digital Health* **4**, e214–e215 (2022).

Supplementary Information

Model-based eye-tracking: a new window to understand individual differences and psychiatric disorders

Qianying Wu[1], Na Yeon Kim[1,2], Ralph Adolphs[1]

[1]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA
[2]Department of Psychology, University of California, Riverside, Riverside, CA, USA

Correspondence

Qianying Wu: qwu@caltech.edu

S1. Conventional eye movement types and their analysis

Saccade & Fixation: Saccades, the most common behavior when we view a scene, are ballistic rapid eye movements from one point (fixation) to another. During saccades, visual perception is suppressed, and thus our perception of the world is smoothly continuous. Fixations are typically brief (around 200 ms) pauses between saccades[1]. For data analysis, saccades and fixations are identified from the timeseries of gaze positions that is measured by an eye tracker, based on the velocity and acceleration of the eye movement. Commonly used metrics include fixation duration, first fixation latency (often from an intertrial fixation cross), saccade amplitude (distance), and saccade velocity.

Microsaccade: Microsaccades are small, involuntary movements that occur spontaneously during fixation. These movements typically occur 1-3 times per second, and lasts 6-30 ms[2]. Microsaccades have been found to be closely associated with covert attention shifts[3]. Similar to saccades, microsaccades are most often defined by thresholds of the velocity, amplitude and duration[4].

Smooth pursuit: Smooth pursuit occurs when the eyes track a moving object and keep the object stabilized on the fovea. Unless highly trained, one cannot make a smooth pursuit without a moving target[5]. In fact, like saccades, it is observed during the rapid eye movement (REM) epochs that accompany dreaming, and has been used as evidence that dreamers can experience conscious visual imagery of moving objects[6]. Smooth pursuit is a clinical marker of schizophrenia. When schizophrenia patients are asked to track a moving target with their eyes, they tend to show lagged eye trajectories following the target, and catch-up saccades immediately afterwards[7].

Area-of-Interest analysis: To establish associations between eye movements and visual stimuli, one could quantify the duration, frequency, and latency of fixations within a contiguous region of the stimulus - an area of interest (AOI). The AOI can be spatially defined (e.g., left vs. right side of the screen) or feature-based (e.g., region of the image where a face is located). However, many feature-based analyses do not lend themselves to AOI analyses because the features are spatially distributed (e.g., the low level visual saliency of an image).

Heatmap & Scanpath visualization: Heatmaps and scanpaths are useful ways to visualize eye movement data (**Figure 1b**). Heatmaps highlight fixation hotspots of an image, such as objects and faces. Scanpaths present a summary of temporally ordered saccades and fixations onto an image. Typically, scanpaths generated from complex images contain about 3-5 fixations per second.

S2. Reliability and validity of eye-tracking research

Test-retest reliability

The variance of any measurement arises from three sources: inter-individual variability, within-individual changes, and random noise. Although many eye-tracking measures are proposed to reflect trait-like characteristics associated with psychiatric features, they are often assessed at a single time point, leaving their test–retest reliability underreported. Short-term test–retest reliability (over intervals of several hours to weeks) evaluates measurement consistency and helps isolate random noise. In contrast, long-term test–retest reliability (over months to years) captures meaningful within-individual changes and informs whether a measure reflects a stable trait or a fluctuating state. Existing eye-tracking metrics exhibit a wide range of reliability - from poor to excellent - across studies[8–11]. Comprehensive assessment of test–retest reliability is essential for determining the robustness of new findings and for establishing the clinical utility of eye-tracking measures as potential diagnostic or intervention markers.

Construct Validity

While computational models provide powerful ways to quantify variables and the relations among them, their interpretation depends on understanding what it is that the variables represent in the first place. Without thorough examination, some of the variables in a model may suffer from poor construct validity (i.e., they do not accurately reflect the theoretical concept they intended to measure), due to noisy measures or confounds at various stages of the experiment. For example, when comparing gaze patterns between younger and older adults, unaccounted influence from poor visual acuity (which is more common in older adults) can confound interpretations of how visual search patterns differ with age. Similarly, when studying children as well as certain clinical groups, the inability to sit still and watch on-screen stimuli is also a significant confound that can obscure conclusions about attention preferences[12,13].

Ecological Validity

Generalizability of findings comes in degrees. Beyond demographic factors such as language, culture, and sex, we often seek conclusions that extend to human cognition and its dysfunction more broadly. One route to broader generalization lies in the diversity of experimental stimuli, which can vary in complexity and representativeness of real-world experience. Historically, most eye-tracking experiments have used static images and revealed robust individual differences in gaze patterns[9,14–19]. However, recent paradigms using movies, immersive virtual reality, or real-world first-person eye-tracking[19–29] have emphasized higher ecological validity – the extent to which findings generalize to the natural environments.

The need for ecological validity depends on the research goal. If the aim is clinical - to find a biomarker for screening and diagnosis, then paradigms that effectively elicit group differences may be more valuable than those maximizing naturalism. For example, smooth pursuit measures in schizophrenia rely on simple moving targets and outperform more naturalistic tasks in distinguishing patients from controls[30,31]. In fact, stimuli that are supposed to be more naturalistic could perform worse if the resulting metric serves as a biomarker: one study showed that fixation duration differences between SZ and controls are much less in the naturalistic environment compared to laboratory tests[32]. Similarly, movies, compared to static images from them, evoke stronger synchrony of eye movements, thereby blurring individual variability[22,33].

In contrast, when the goal is to understand underlying mechanisms or etiology, ecological validity becomes crucial. This echoes Brunswik's notion of "representative design," emphasizing that experimental stimuli and tasks should mirror the real-world contexts in which psychological processes occur[34,35]. While some of the findings from static image viewing may well generalize to the real world (e.g., see an example of such success in ref[36]) for some populations, others may not - especially when studying social attention. Real-time social interactions involve eye contact, joint attention, and signaling of intentions or desires, and gaze patterns observed in static face viewing may not generalize to dynamic, interactive situations (e.g., talking faces or mutual gaze)[37,38]. Indeed, gaze-to-face behaviors can differ markedly between interactive and non-interactive contexts[37,39]. The highly context-dependent nature of social attention could lead to conflicting results observed between lab and naturalistic environments. For instance, a study found that individuals with higher social anxiety scores show reduced fixations at persons closer in distance, yet the effect only existed in the real life condition but not the in-lab video-watching condition[40].

Another situation when ecological validity matters is the study of goal-directed attention allocation in everyday life activities, such as driving, navigation, and shopping[41–43]. In these behaviors, attention is guided primarily by top-down control rather than bottom-up visual saliency, and eye movements reflect how individuals flexibly allocate attention to gather information, coordinate actions, reduce uncertainty, and maximize reward[44–46]. Therefore, proximity to real life situations is crucial for capturing the cognitive constraints that shape such behaviors, including time pressure, cognitive load, and motor coordination. These task-oriented gaze patterns have also been modeled under various computational frameworks[47–50], and are promising tools for examining dysfunctions in executive functioning, reward processing, and action planning in the clinical population[41,51].

S3. Camera-based eye-tracking

New technologies use camera input from smartphones, tablets, or glasses offering scalable alternatives to traditional infrared-based desktop systems[25,38,52–57]. These approaches apply computer vision models to detect face and the eye regions, and estimate gaze locations through pre-trained algorithms. Compared to the standard desktop-based, infrared eye-trackers (e.g., Tobii, EyeLink), which achieve high spatial accuracy (less than 1° error) and sampling rates up to 1200 Hz, current camera-based methods generally have limited spatial and temporal (15-50 Hz) resolution. However, gaze estimation has been improving with deep learning-based models, which to date have achieved spatial accuracies as low as 2.4° and precision of 0.47° visual angle[52]. Notably, these methods have been validated in psychology research: a recent paper demonstrated sufficient accuracy of a smartphone-based eye-tracking tool for analyzing gaze on specific features (e.g., human face, body area) of YouTube videos and replicated well-established findings on atypical social attention in autism[25].

Wearable eye-tracking glasses further expand these possibilities by enabling the study of real-world social interactions, such as making and sustaining eye contact. A recent model, the Pupil Labs Neon, does not require any calibration, delivers a spatial resolution of 1.3° visual angle at 200Hz, and collects audio stream and head movement data simultaneously[58,59]. As models continue to improve, camera-based eye-tracking holds great promise for broadening access to large-scale, high-quality gaze data collection across diverse populations.

Reference

1. Bischof, W. F., Anderson, N. C. & Kingstone, A. Temporal Methods for Eye Movement Analysis. in *Eye Movement Research* 407–448 (Springer, Cham, 2019). doi:10.1007/978-3-030-20085-5_10.
2. Gu, Q. *et al.* Microsaccades reflect attention shifts: a mini review of 20 years of microsaccade research. *Front. Psychol.* **15**, (2024).
3. Engbert, R., Mergenthaler, K., Sinn, P. & Pikovsky, A. An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences* **108**, E765–E770 (2011).
4. Hauperich, A.-K., Young, L. K. & Smithson, H. E. What makes a microsaccade? A review of 70 years of research prompts a new detection method. *J Eye Mov Res* **12**, 10.16910/jemr.12.6.13.
5. Purves, D. *et al.* Types of Eye Movements and Their Functions. in *Neuroscience. 2nd edition* (Sinauer Associates, 2001).
6. LaBerge, S., Baird, B. & Zimbardo, P. G. Smooth tracking of visual targets distinguishes lucid REM sleep dreaming and waking perception from imagination. *Nat Commun* **9**, 3298 (2018).
7. Morita, K., Miura, K., Kasai, K. & Hashimoto, R. Eye movement characteristics in schizophrenia: A recent update with clinical implications. *Neuropsychopharmacol Rep* **40**, 2–9 (2019).
8. Recker, L. & Poth, C. H. Test–retest reliability of eye tracking measures in a computerized Trail Making Test. *J Vis* **23**, 15 (2023).
9. de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S. & Gegenfurtner, K. R. Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences* **116**, 11687–11692 (2019).

10. Bargary, G. *et al.* Individual differences in human eye movements: An oculomotor signature? *Vision Res* **141**, 157–169 (2017).

11. Ettinger, U. *et al.* Reliability of smooth pursuit, fixation and saccadic eye movements. *Psychophysiology* **40**, 620–628 (2003).

12. Wass, S. V., Forssman, L. & Leppänen, J. Robustness and Precision: How Data Quality May Influence Key Dependent Variables in Infant Eye-Tracker Analyses. *Infancy* **19**, 427–460 (2014).

13. Wang, Y., Koch, M., Bâce, M., Weiskopf, D. & Bulling, A. Impact of Gaze Uncertainty on AOIs in Information Visualisations. in *2022 Symposium on Eye Tracking Research and Applications* 1–6 (Association for Computing Machinery, New York, NY, USA, 2022). doi:10.1145/3517031.3531166.

14. Wang, S. *et al.* Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron* **88**, 604–616 (2015).

15. Adámek, P. *et al.* The Gaze of Schizophrenia Patients Captured by Bottom-up Saliency. *Schizophr* **10**, 1–13 (2024).

16. Griffin, J. W. *et al.* Spatiotemporal Eye Movement Dynamics Reveal Altered Face Prioritization in Early Visual Processing Among Autistic Children. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* https://doi.org/10.1016/j.bpsc.2024.08.017 (2024) doi:10.1016/j.bpsc.2024.08.017.

17. Benson, P. J. *et al.* Simple Viewing Tests Can Detect Eye Movement Abnormalities That Distinguish Schizophrenia Cases from Controls with Exceptional Accuracy. *Biological Psychiatry* **72**, 716–724 (2012).

18. Kennedy, D. P. *et al.* Genetic Influence on Eye Movements to Complex Scenes at Short Timescales. *Current Biology* **27**, 3554-3560.e3 (2017).

19. Shic, F. *et al.* The Autism Biomarkers Consortium for Clinical Trials: evaluation of a battery of candidate eye-tracking biomarkers for use in autism clinical trials. *Molecular Autism* **13**, 15 (2022).

20. Keles, U. *et al.* Atypical gaze patterns in autistic adults are heterogeneous across but reliable within individuals. *Molecular Autism* **13**, 39 (2022).

21. Campbell, D. J., Shic, F., Macari, S. & Chawarska, K. Gaze Response to Dyadic Bids at 2 Years Related to Outcomes at 3 Years in Autism Spectrum Disorders: A Subtyping Analysis. *J Autism Dev Disord* **44**, 431–442 (2014).

22. Dorr, M., Martinetz, T., Gegenfurtner, K. R. & Barth, E. Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision* **10**, 28–28 (2010).

23. Jones, W. *et al.* Development and Replication of Objective Measurements of Social Visual Engagement to Aid in Early Diagnosis and Assessment of Autism. *JAMA Network Open* **6**, e2330145 (2023).

24. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344 (2017).

25. Kim, N. Y. *et al.* Smartphone-based gaze estimation for in-home autism research. *Autism Research* **17**, 1140–1148 (2024).

26. Haskins, A. J., Mentch, J., Botch, T. L. & Robertson, C. E. Active vision in immersive, 360° real-world environments. *Sci Rep* **10**, 14304 (2020).

27. Zheng, Z. *et al.* Diagnosing and tracking depression based on eye movement in response to virtual reality. *Front Psychiatry* **15**, 1280935 (2024).

28. Mendez-Encinas, D., Sujar, A., Bayona, S. & Delgado-Gomez, D. Attention and impulsivity assessment using virtual reality games. *Sci Rep* **13**, 13689 (2023).

29. Wiebe, A. *et al.* Virtual reality-assisted prediction of adult ADHD based on eye tracking, EEG, actigraphy and behavioral indices: a machine learning analysis of independent training and test samples. *Transl Psychiatry* **14**, 508 (2024).

30. O’Driscoll, G. A. & Callahan, B. L. Smooth pursuit in schizophrenia: A meta-analytic review of research since 1993. *Brain and Cognition* **68**, 359–370 (2008).

31. Athanasopoulos, F., Saprikis, O.-V., Margeli, M., Klein, C. & Smyrnis, N. Towards Clinically Relevant Oculomotor Biomarkers in Early Schizophrenia. *Front. Behav. Neurosci.* **15**, (2021).

32. Dowiasch, S. *et al.* Eye movements of patients with schizophrenia in a natural environment. *Eur Arch Psychiatry Clin Neurosci* **266**, 43–54 (2016).

33. Loschky, L. C., Larson, A. M., Magliano, J. P. & Smith, T. J. What Would Jaws Do? The Tyranny of Film and the Relationship between Gaze and Higher-Level Narrative Film Comprehension. *PLOS ONE* **10**, e0142474 (2015).

34. Brunswik, E. Representative design and probabilistic theory in a functional psychology. *Psychological Review* **62**, 193–217 (1955).

35. Holleman, G. A., Hooge, I. T. C., Kemner, C. & Hessels, R. S. The ‘Real-World Approach’ and Its Problems: A Critique of the Term Ecological Validity. *Front. Psychol.* **11**, (2020).

36. Peterson, M. F., Lin, J., Zaun, I. & Kanwisher, N. Individual differences in face-looking behavior generalize from the lab to the world. *Journal of Vision* **16**, 12 (2016).

37. Hessels, R. S. How does gaze to faces support face-to-face interaction? A review and perspective. *Psychon Bull Rev* **27**, 856–881 (2020).

38. Valtakari, N. V. *et al.* Eye tracking in human interaction: Possibilities and limitations. *Behav Res* **53**, 1592–1608 (2021).

39. Gado, S., Teigeler, J., Kümpel, K., Schindler, M. & Gamer, M. The effect of social anxiety on social attention in naturalistic situations. *Anxiety, Stress, & Coping* **38**, 326–342 (2025).

40. Rubo, M., Huestegge, L. & Gamer, M. Social anxiety modulates visual exploration in real life - but not in the laboratory. *Br J Psychol* **111**, 233–245 (2020).

41. Wolf, A. & Ueda, K. Contribution of Eye-Tracking to Study Cognitive Impairments Among Clinical Populations. *Front. Psychol.* **12**, (2021).

42. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. *Trends in Cognitive Sciences* **9**, 188–194 (2005).

43. Hayhoe, M. & Ballard, D. Modeling Task Control of Eye Movements. *Current Biology* **24**, R622–R628 (2014).

44. Henderson, J. M., Brockmole, J. R., Castelhano, M. S. & Mack, M. Visual saliency does not account for eye movements during visual search in real-world scenes. in *Eye movements: A window on mind and brain* 537–562 (Elsevier, Amsterdam, Netherlands, 2007). doi:10.1016/B978-008044980-7/50027-6.

45. Rothkopf, C. A., Ballard, D. H. & Hayhoe, M. M. Task and context determine where you look. *Journal of Vision* **7**, 16 (2016).

46. Shimojo, S., Simion, C., Shimojo, E. & Scheier, C. Gaze bias both reflects and influences preference. *Nat Neurosci* **6**, 1317–1322 (2003).

47. Krajbich, I., Armel, C. & Rangel, A. Visual fixations and the computation and comparison of value in simple choice. *Nat Neurosci* **13**, 1292–1298 (2010).

48. Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R. & Mohr, P. N. C. Gaze bias differences capture individual choice behaviour. *Nat Hum Behav* **3**, 625–635 (2019).

49. Eckstein, M. P., Drescher, B. A. & Shimozaki, S. S. Attentional Cues in Real Scenes, Saccadic Targeting, and Bayesian Priors. *Psychol Sci* **17**, 973–980 (2006).

50. Zhu, S., Lakshminarasimhan, K. J., Arfaei, N. & Angelaki, D. E. Eye movements reveal spatiotemporal dynamics of visually-informed planning in navigation. *eLife* **11**, e73097 (2022).

51. Spering, M. Eye Movements as a Window into Decision-Making. *Annual Review of Vision Science* **8**, 427–448 (2022).

52. Saxena, S., Fink, L. K. & Lange, E. B. Deep learning models for webcam eye tracking in online experiments. *Behav Res* **56**, 3487–3503 (2024).

53. Werchan, D. M., Thomason, M. E. & Brito, N. H. OWLET: An automated, open-source method for infant gaze tracking using smartphone and webcam recordings. *Behav Res* **55**, 3149–3163 (2023).

54. Erel, Y. *et al.* iCatcher+: Robust and Automated Annotation of Infants' and Young Children's Gaze Behavior From Videos Collected in Laboratory, Field, and Online Studies. *Advances in Methods and Practices in Psychological Science* **6**, 25152459221147250 (2023).

55. Chang, Z. *et al.* Computational Methods to Measure Patterns of Gaze in Toddlers With Autism Spectrum Disorder. *JAMA Pediatrics* **175**, 827–836 (2021).

56. Valliappan, N. *et al.* Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nat Commun* **11**, 4553 (2020).

57. Cheng, Y., Wang, H., Bao, Y. & Lu, F. Appearance-based Gaze Estimation With Deep Learning: A Review and Benchmark. Preprint at https://doi.org/10.48550/arXiv.2104.12668 (2024).

58. Kaplan, B. E., Martinez, E. & Yu, C. Using Head-Mounted Eye Tracking to Examine Infant Face Looking During Naturalistic Freeplay. *Proceedings of the Annual Meeting of the Cognitive Science Society* **47**, (2025).

59. Hessels, R. S. *et al.* Eye contact avoidance in crowds: A large wearable eye-tracking study. *Atten Percept Psychophys* **84**, 2623–2640 (2022).