A Survey of Incorporating Gaze Data into Natural Language Processing Models and Applications

Cengiz Acartürk[1], Burcu Can[2], Melike Çağlayan[1], Jamal Abdul Nasir[3], Çağrı Çöltekin[4]

[1]Centre for Cognitive Science, Jagiellonian University, [2]Department of Computing Science and Mathematics, University of Stirling, [3]School of Computer Science, University of Galway

[4]Department of Linguistics, University of Tübingen

cengiz.acarturk@uj.edu.pl, burcu.can@stir.ac.uk, melike.caglayan@uj.edu.pl, jamal.nasir@universityofgalway.ie, cagri.coeltekin@uni-tuebingen.de

Abstract

This study presents a survey of research integrating eye-tracking (gaze) data into Language Models (LMs) as a means of cognitively grounding NLP models and applications in human reading behavior. Although contemporary LMs excel at learning statistical patterns from text, they fundamentally lack human-like reading and comprehension capabilities. Incorporating gaze data may offer a window into cognitive processing, yet its impact on LMs remains underexplored. Addressing a persistent bottleneck, namely, the high cost and limited scale of laboratory eye-tracking, we propose a roadmap consisting of three streams of research for advancing this novel research domain: (1) developing cognitive multimodal corpora, (2) leveraging generative models for gaze synthesis to overcome the data bottleneck caused by the high costs of human eye-tracking, and (3) training LMs with gaze-guided attention mechanisms and input augmentation. Furthermore, we illustrate practical applications in readability assessment, educational analytics, and assistive communication, demonstrating how gaze-informed models can enable adaptive technologies. Finally, we critically examine ongoing challenges, including the lack of data standardization, the misalignment between human and machine language processing, and the urgent ethical imperative for privacy-preserving architectures to protect sensitive biometric gaze data, motivating privacy-aware data practices and model designs for scalable deployment.

Keywords: Gaze data, Language Models, Human Reading Behavior

1. Introduction

Natural Language Processing (NLP) has achieved remarkable success, driven largely by deep learning architectures that learn statistical patterns from massive text corpora (language models, henceforth LMs). However, LMs often lack the cognitive grounding that characterizes human language processing and comprehension. Eye-tracking data, specifically the measurement of fixations and saccades, provide a psychophysiological window into these cognitive processes. Common gaze features used in NLP include fixation duration (processing effort), first-pass reading time (early processing), total reading time (integration difficulty), regression rate (reanalysis), and skipping probability (predictability). These features provide measurable indicators of cognitive processing during reading and are widely used in computational models.

The integration of gaze data into NLP models and applications offers a pathway to bridge the gap between statistical probability and human-like reading and comprehension by informing models about which words might carry the most information, and how attention and cognitive load might change during reading from a human-reader perspective (cf., Cognitively Inspired NLP, Mishra et al. 2017; cognitive signals, Hollenstein et al. 2020a). This provides the development of a novel research domain that bridges NLP and cognitive sciences, attracting current research and proceeding towards establishing a community of researchers (Acartürk et al., 2025). To give a concrete example to the studies in this context, gaze data may bridge a text-specific property, such as sentence complexity, to its human-reader counterpart, namely, reading difficulty.

Recently, the research on developing computational models of eye-movement during reading faced a significant data bottleneck, as collecting gaze data involves high cost and time effort, also requiring specialized hardware and controlled laboratory environments. These challenges have limited the use of gaze data in NLP to relatively small-scale studies. However, a recent shift has been moving the eye-tracking field from small datasets to standardized multimodal and multilingual corpora by incorporating synthetic gaze generation.

This survey reviews the current role of gaze data in NLP research and outlines a structured research direction through a three-stream roadmap (see Figure 1): (1) Developing cognitive multimodal corpora, (2) enriching cognitive multimodal corpora by gaze synthesis, and (3) training LMs for gaze guidance embedded in enriched cognitive multimodal corpora. We also address methodological and ethical considerations for large-scale collection, synthesis, and deployment due to the personal and potentially sensitive nature of gaze data.

Proceedings of The Second International Workshop onEye-Tracking Resources and Evaluation for Human-Aligned NLP, pages 64–76
May 12, 2026.

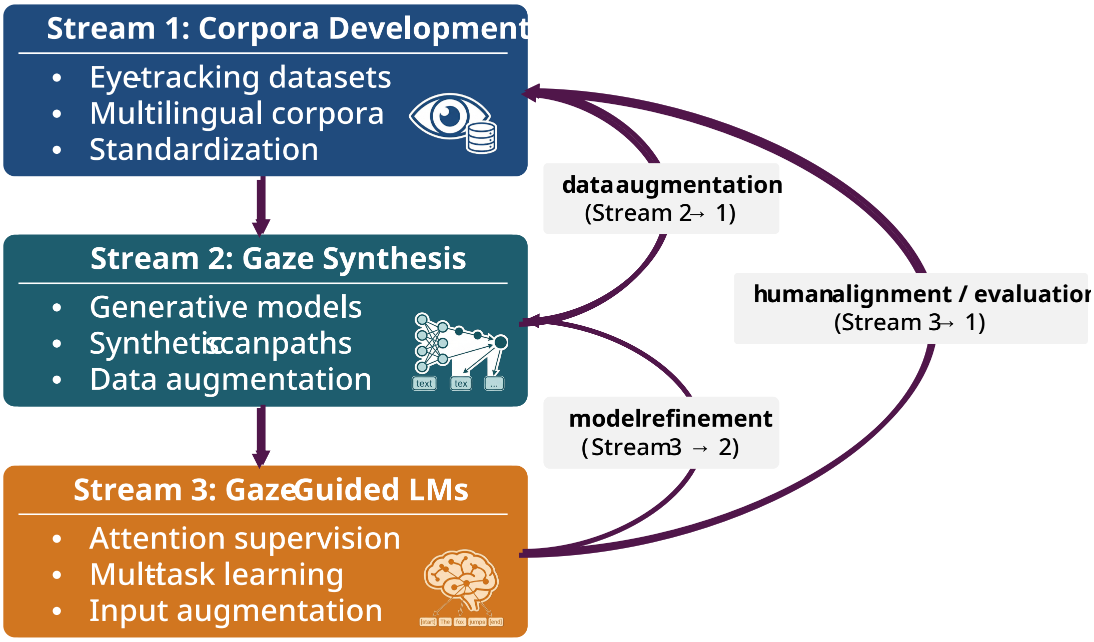

Figure 1: Roadmap for gaze-integrated NLP. The three streams for corpora, gaze synthesis, and gaze-guided models can be developed in parallel, with feedback loops that support continuous improvement and data augmentation.

2. A Roadmap for Gaze for NLP

This section introduces the roadmap in three streams of research that can proceed in parallel with minor dependencies. The first stream presents the studies targeting the development of cognitive multimodal corpora. Given their limited size, due to challenges in developing multimodal corpora, the next stream addresses synthesizing gaze data to enrich multimodal corpora and the studies conducted in this direction. The next stream focuses on training LMs on enriched cognitive multimodal corpora through gaze guidance.

2.1. Developing Cognitive Multimodal Corpora

Developing corpora has been the very first step for conducting quantitative linguistic research and developing practical NLP applications for many decades. The scope of corpus development has been expanded for the past decade in at least two directions: The development of text-only corpora and the development of multimodal cognitive corpora, both for training LMs. Due to historical reasons, several terminological ambiguities have emerged. One is the use of the term “multimodal”, which has been used in NLP studies to refer to language resources including linguistic and nonlinguistic content, such as text and images. Another use of the term refers to the method of measurement and the resulting data, such as gaze and brain imaging data. To resolve the terminological ambiguity in the field, we propose a distinction between extrinsic multimodality (content-multimodal corpora), which concerns the diverse formats of information representation (e.g., text paired with images or video), and intrinsic multimodality (cognitive multimodal corpora), which concerns the diverse channels of human cognitive processing (e.g., simultaneous recording of gaze, EEG, and fMRI while reading text). This survey specifically addresses the latter, using the term cognitive multimodal corpora, by leveraging intrinsic cognitive signals to ground language representations in human processing patterns. While extrinsic multimodal systems learn to relate objects and concepts across different external media, intrinsic multimodal systems learn to model the observable signals obtained from the reader.

Several cognitive multimodal corpora have been released during the past decade. The datasets, such as ZuCo 2.0 Zurich Cognitive Language Processing Corpus (Hollenstein et al., 2020b) and GECO Ghent Eye-Tracking Corpus (Cop et al., 2017) provide standardized benchmarks. For instance, ZuCo 2.0, provides simultaneous eye-tracking and EEG recordings during both natural reading and specific annotation tasks of 739 English sentences read by 18 participants. Similarly, GECO consists of gaze data from monolingual English speakers and Dutch-English L2 learners reading an entire novel, providing a continuous narrative context often missing in sentence-level datasets.

While early work focused primarily on English, the field has rapidly expanded into diverse languages and specialized domains to create standardized benchmarks. Recent efforts include the Copenhagen Corpus (CopCo), which provides the first eye-tracking-while-reading corpus for the Danish language, consisting of 1,832 sentences and nearly 35,000 tokens (Hollenstein et al., 2022). Similarly, the PoTeC, Potsdam Textbook Corpus (Jakobi et al., 2025a), is a German naturalistic resource comprising data from 75 participants reading scientific texts. It employs a factorial design contrasting expert and novice readers to investigate how domain expertise and prior knowledge influence cognitive processing during reading. Recently, GAZE4HATE dataset expands the scope of reading stimuli from common text to instances of annotated hate-speech segments (Alacam et al., 2024).

To address the historical lack of uniformity across laboratories, the MultiplEYE initiative has been establishing a large-scale, multilingual corpus with consistent recording protocols (Jakobi et al., 2025b) Novel datasets like CoLAGaze (Bondar et al., 2025) provide the first broad-coverage eye-tracking corpus on grammatical and ungrammatical sentences, aligned to the CoLA benchmark, enabling direct comparison between model grammaticality judgments and human processing patterns. Furthermore, WebQAmGaze (Ribeiro et al., 2023) addresses accessibility, utilizing webcam-based eye tracking from 600 participants across four languages (English, German, Spanish, and Turkish), demonstrating that low-cost, scalable data collection is feasible without specialized hardware. InteRead (Zermiani et al., 2024), with 50 participants, annotates interruptions and resumption lags during reading, addressing the ecological validity of learning environments. Additional multilingual resources continue to expand the range of cognitive multimodal corpora. For example, the FETA corpus provides French eye-tracking data collected from 46 readers across general, medical, and clinical texts presented in both original and manually simplified versions, with multiple word-level gaze features available for analysis (Ivchenko and Grabar, 2025). To provide a clearer overview of existing resources, Table 1 summarizes representative cognitive multimodal corpora.

The available cognitive multimodal corpora, as exemplified in this section, provides valuable resources for developing LMs by incorporating gaze data. On the other hand, a major challenge in the field is the limited size and scope of the available datasets, mostly consisting of a few dozen participants and a few thousand words, due to the high cost and time effort required in gaze data collection. This scarcity has necessitated a shift in research focus for the past decade, seeking an answer to the following question: Is it possible to synthesize gaze data to predict human gaze patterns on unseen text, so as to scale up training sets for gaze-augmented models? The studies in this direction comprise the second stream of the research roadmap, presented below.

2.2. Gaze Synthesis to Enrich Cognitive Multimodal Corpora

The predictive models of eye movements during reading emerged in the beginning of the century, established on experimental findings obtained in previous research. The empirical research largely addressed the factors impacting eye-movement patterns on text, such as corpus frequency of lexical items, word length, and predictability of words in a sentential context. A set of computational cognitive models aimed to parameterize these factors to predict when and where to move the eyes during reading (E-Z Reader, Reichle et al. 1998; SWIFT Saccade-generation with inhibition by foveal targets, Engbert et al. 2002; Glenmore, Reilly and Radach 2003; OB1-reader, Snell et al. 2018; see Acartürk 2025 for a recent review). For the past several decades, statistical regularities on eye movements during reading have attracted research interest from a modeling point of view, still being an ongoing debate (Kimchi and Siegelman, 2026).

The scope of early modeling in the cognitive computational models has been limited to rule-based, algorithmic approaches while some employed connectionist architectures or parameter optimization. Recent models have involved explicit machine-learning approaches, also targeting synthesized gaze patterns. For instance, models like ScanTextGAN and the hybrid text saliency models proposed by Khurana et al. (2023) indicate that synthetic scanpaths can be used to supervise NLP models. Specifically, they were able to improve the accuracy of multiple NLP tasks, such as sentiment analysis, named entity recognition, relation extraction, and grammatical error detection, using synthetic scanpaths produced by the model, on a variety of datasets. ScanTextGAN presents a generative approach designed to synthesize human-like scanpaths (the sequence of fixations and saccades) over text, also aiming to replicate the temporal sequence of reading, including regressions (looking back) and skips. The model was trained on real eye-tracking data to learn the probability distribution of eye movements during reading. Hybrid and cognitive text saliency models aim to predict aggregate gaze metrics (like total fixation duration) rather than full scanpaths, yet combining traditional cognitive reading models with data-driven gaze supervision. This approach reduces the full reliance on black-box neural models, by integrating established psycholinguistic features (e.g., word length, frequency) with neural attention. Empirical findings

Dataset	Language	Participants	Size
ZuCo 2.0	English	18	739 Sentences
GECO	English, Dutch (L2)	30	5,000 Sentences
CopCo	Danish	22	1,832 Sentences
PoTeC	German	75	12 Scientific Texts
FETA	French	46	32 → 179 Sentences
WebQAm Gaze	English, German, Spanish, Turkish	600	XQuAD & MECO texts
Inte Read	English	50	28 Pages
CoLA Gaze	English	42	306 Sentences
Multipl EYE	Multilingual (30+ languages)	100+*	10 Natural Texts
GAZE4HATE	English	43	90 items

Table 1: Overview of representative cognitive multimodal corpora with eye-tracking data (*per lab, ongoing)

suggest that the field will increasingly pivot toward gaze synthesis in the near future, relying on the expectation that the accuracy of synthesized gaze will further and further approximate real gaze data.

Alongside ScanTextGAN, several architectures have emerged for gaze synthesis. Eyettention (Deng et al., 2023) is such a gaze-synthesis model, one that processes linguistic tokens and chronological fixation sequences simultaneously using cross-sequence attention. By modeling the complex temporal dynamics of reading, Eyettention performs scanpath prediction across multiple languages and datasets. Another model is ScanEZ (Sood et al., 2025), which integrates traditional cognitive models with self-supervised learning to produce spatiotemporally realistic scanpaths. This hybrid approach demonstrates that incorporating traditional, computational cognitive models that address psycholinguistic factors can improve generalization beyond purely data-driven methods.

These synthesis approaches differ among each other in both their representational assumptions and architectural design. Sequence-based models conceptualize scanpaths as ordered discrete events (i.e., fixations and saccades), typically modeled with autoregressive or recurrent frameworks. Complementary work has also investigated the linguistic determinants of saccade targeting, showing that, beyond word length and frequency, contextual meaning, prior fixation history, and saccade distance contribute to predicting forward and backward eye movements during naturalistic reading (Rego et al., 2025). In contrast, trajectory-based models treat gaze as a continuous spatiotemporal signal and often employ probabilistic or neural dynamical systems to capture smooth transitions in gaze coordinates. Conditioning strategies also vary: some models incorporate explicit linguistic features (e.g., readability indices, syntactic complexity, or lexical frequency), whereas others learn end-to-end mappings from token-level embeddings directly to gaze trajectories without handcrafted features.

While generating synthetic gaze data offers a theoretical workaround to the high cost of collecting human eye-tracking data, it is crucial to recognize significant limitations inherent in relying on synthesized data. Currently, it remains unclear how well synthetic gaze truly captures the complexity of human reading behavior. A primary shortcoming of synthetic gaze data is the reduction of natural variability. Because generative models are trained to optimize for widespread statistical regularities, they risk ignoring the rich, idiosyncratic variations present in human reading patterns. Another weakness related to the existing models is the lack of assessment methodologies in the recent work. In general, the evaluation of the relationship between machine behavior and human behavior is scarce in the context of incorporating gaze data into LMs, except for a few studies in this direction, such as Ikhwantri et al. (2023). Aggregate behavioral statistics (e.g., mean fixation duration, saccade length, regression rate) or predictive measures (e.g., next-fixation accuracy) are generally used to measure synthetic gaze. Although informative, these measures mainly capture superficial likeness and may not indicate whether the generated scanpaths reflect the finer-grained linguistic processes in human language processing (e.g., sensitivity to syntactic boundaries, semantic ambiguity, or surprisal). As a result, matching global distributions does not guarantee that synthetic gaze preserves the cognitively compatible signal required for modeling. Specifically, current synthetic models generally fail to model individual differences. Reader-specific traits, such as working memory capacity, domain expertise, second-language proficiency, or reading disorders, profoundly influence gaze behavior. By heavily relying on aggregated training sets, generative models run the risk of biasing NLP applications toward the “average gazer.” This homogenization means that models might ignore outlier behaviors or non-standard reading strategies, limiting the ecological validity and inclusivity of applications like assistive communication or personalized education.

In summary, the methodological issue of trying to determine that the synthesized scanpaths do not store task-relevant information content, but are just an approximation of statistical regularities, is still unresolved. It is not clear that existing generative models reproduce systematic variability with regard to individual variability, task requirements, and processing linguistically complex or ambiguous stimuli. Synthetic gaze augmentation effectiveness is determined by whether models are able to encode both the statistical patterns of eye movements and the structured modulation of gaze based on textual properties, reader characteristics, and contextual constraints.

Consequently, while gaze synthesis provides an alternative avenue for data augmentation, its utility must be critically weighed against these conceptual and methodological shortcomings. Overreliance on synthetic datasets without addressing the loss of individual variability and nuanced cognitive processing poses a major risk. Acknowledging these limitations, the next challenge lies in architectural integration: How can both empirical and (cautiously applied) synthetic cognitive signals be effectively injected into neural NLP models? The studies that aim to find answers to these questions comprise the third and final stream of the research roadmap, presented below.

2.3. Gaze Guidance for Training LMs with Enriched Cognitive Multimodal Corpora

The previous studies on incorporating gaze data into LMs have mainly converged on two strategies: augmenting the input representation (concatenation) and supervising the internal computation (attention mechanisms). The relationship between gaze and visual attention is basically an operational assumption rather than being an observational fact. The eye movements during reading indicate visual attention while the two do not necessarily couple momentarily (Posner et al., 1980). Moreover, aligning a model's attention mechanism with human visual attention is also a technically sophisticated integration approach. In standard Transformer or RNN-based architectures, "attention" is a learned weight distribution that determines which input words the model should focus on. A feasible approach is to modify the loss function of the neural network to minimize the distance between the model's attention weights and human fixation distributions (or synthetic equivalents) instead of letting the model learn attention solely from the target task (e.g., translation). This approach assumes that human gaze is a proxy for attention. Recent research shows that integrating gaze features into the attention mechanisms of neural networks can improve performance in tasks like paraphrase generation and sentence compression (Khurana et al., 2023). Such supervised learning of attention mechanisms combined with gaze data provide a practical benefit, as once training is complete, it does not require gaze input during inference. It allows building cognitively enriched attention mechanisms, which have signals from both raw textual data and cognitive data.

Another approach for informing NLP systems with gaze data relies on joint, multi-task models that learn gaze behaviour alongside the target NLP task. In this approach, the model is trained to perform two tasks simultaneously: a primary task consisting of a specific NLP goal (e.g., sentence compression, sentiment analysis), and an auxiliary task aiming to predict gaze metrics (e.g., total fixation duration) for the input tokens. By sharing the underlying encoder layers between these two tasks, the model learns a generalized text representation that encodes both semantic meaning and cognitive processing effort. This joint-modeling approach ensures that gaze data regularizes the model, preventing it from overfitting to superficial statistical patterns in the text. Sood et al.'s (2020b) joint modeling approach—training the NLP model to solve the task and predict human gaze simultaneously— outperformed state-of-the-art baselines. Similarly, Mathias et al. (2020) introduce a model where gaze behaviour is learned along with essay grading, and the results show that modeling gaze behaviour along with essay grading provides statistically significant performance gains for the task.

A similar approach is to concatenate gaze vectors with word embeddings to enrich the input representation. Gaze measures can also be normalized and concatenated directly with word embeddings (e.g., BERT or GloVe vectors) to form the input to the network. However, this method increases the dimensionality of the input and requires the gaze data to be available at test time (unless replaced by synthetic features), making it less flexible than the attention supervision methods. On the contrary to the previous approach, where the attention mechanisms are trained along with the gaze features, this approach requires gaze features during inference time to be available, which makes the approach less practical compared to the previous approach.

Empirical evaluations of these approaches have yielded mixed but instructive results. Barrett et al. (2018) showed that regularizing recurrent models with human attention from eye-tracking corpora improved performance across sentiment analysis, grammatical error detection, and abusive language detection, suggesting that human gaze provides inductive biases that help models generalize. Similarly, Wang B. et al. (2024) investigated whether integrating gaze signals into BERT during pretraining or fine-tuning enhances performance, reporting gains on several tasks. Hollenstein and Zhang (2019) demonstrated consistent improvements from gaze and EEG augmentation for named entity recognition, relation classification, and sentiment analysis across multiple datasets. However, Sood et al. (2020a) found that while model attention can be shaped to resemble human attention through supervision, the correlation between attention similarity and task performance is architecture-dependent: LSTM and CNN attention aligned better with human fixations and correlated with performance, whereas XLNet showed weaker alignment despite high task accuracy. This suggests that the concept of attention supervision can be best implemented when the internal process of the model can be compatible with the human-like sequential processing. The major outstanding problem is to enable type-level generalization, i.e. learners to use the gaze patterns trained on training texts without having to provide eye-tracking information at testing time (Hollenstein and Zhang, 2019).

In summary, training LMs with enriched cognitive multimodal corpora seems to be a promising research direction for effectively incorporating gaze data. Below, we present research domains that might extend the frontiers of the field through applications.

3. Applications

This section exemplifies several frontiers that the research on incorporating gaze data into LMs has been expanding

3.1. Text Complexity and Reading Difficulty

Text complexity and reading difficulty are two concepts that constitute the two piers of the bridge between NLP and cognitive sciences. The link between gaze behavior and text difficulty has been well established in psycholinguistics, but translating this link into computational models is not straightforward. Text difficulty is not a single construct; it depends on lexical properties, syntactic structures, discourse relations, and reader-dependent variables such as reading skill and prior knowledge. These factors interact and produce variable eye-movement patterns across readers and tasks.

Traditional readability formulas mainly use surface features such as word length and sentence length. These features provide an incomplete proxy for processing effort and do not directly reflect cognitive operations during reading, such as lexical access, syntactic parsing, and semantic and discourse integration. Eye-tracking studies show that fixation duration, regression rate, and skipping probability are influenced by linguistic predictors like predictability, and syntactic ambiguity. As a result, surface readability scores often fail to capture the variation observed in human gaze data. For computational modeling, this creates a mismatch between input representations and gaze signals. Gaze data are temporally structured and respond to local linguistic properties at the word and sentence level, while many traditional NLP models use global or static text features to represent difficulty. A more suitable approach is to incorporate psycholinguistically motivated predictors, such as surprisal, entropy, and syntactic dependency measures, alongside contextual embeddings. Recent work further shows that surprisal estimates from LLMs predict several eye-tracking measures, including first fixation duration and gaze duration, reproducing well-established reading-time effects across languages and datasets (Alves, 2025). This alignment can improve the mapping between text features and gaze patterns, supporting more accurate modeling of reading behavior. Overall, gaze-augmented methods aim to assess text difficulty by measuring actual processing effort during reading. Regression count indexes reanalysis caused by comprehension problems or structural ambiguity. First-pass reading time reflects early lexical access and syntactic integration, whereas total reading time captures later reanalysis and cumulative processing. These measures can diverge: a frequent word may show short first-pass time but longer total time when it appears in syntactically demanding contexts, indicating interaction between lexical and syntactic processing that surface features miss.

Nevertheless, applying these findings to readability systems raises practical issues. Gaze data shows high variability across readers due to differences in skill, working memory, domain knowledge, and reading strategy. Aggregating at the type level can mask this variability, while token-level modeling requires large datasets for stable estimates. In addition, the link between gaze and comprehension depends on task demands (e.g., skimming vs. careful reading) and genre, which limits model generalization across corpora. Recent work has started to operationalize these findings. Studies that combine gaze features with linguistic and psycholinguistic variables in multi-task learning models report improved readability prediction in English (González-Garduño and Søgaard, 2017) and in morphologically rich languages such as Arabic (Baazeem et al., 2025). Extending this line of work to lower-resource languages, Hodivoianu et al. (2025) introduce the first Romanian eye-tracking dataset for reading and show that both feature-based models and fine-tuned Romanian BERT architectures can predict total reading time at the word level, with applications to lexical simplification and interactive readability support. Transformer-based language models have been shown to better account for human reading effort measures than RNNs, including self-paced reading times and neural activity during sentence processing (Merkx and Frank, 2021). These findings suggest that attention-based architectures can capture aspects of human language processing, although their fit to behavioral data depends on the type of measures and the experimental setting.

3.2. Educational Analytics

A promising field of research that benefits from incorporating gaze data into LMs has been educational analytics. For instance, foreign language learning and its evaluation by automated essay scoring are the two domains that can be bridged by incorporating gaze data into LMs. In educational settings, gaze data has the potential to serve as a measure of learning Sharma et al. (2020); John et al. (2025). Although gaze is an indirect measure of learner engagement and comprehension difficulty (Hutt et al., 2024; Turčáni et al., 2024), specific gaze measures, such as regressive saccades, may correlate with various levels of text complexity, such as lexical- and syntactic-level complexity (Turčáni et al., 2024), also indicating challenges in language learning. An emerging application leverages gaze data to infer open-ended reading goals and information-seeking intentions from eye movements, without explicit user input (Hadar et al., 2025; Shubi et al., 2025). Multimodal LLMs trained on large-scale eye-tracking data have achieved promising success in predicting whether a reader is engaged in comprehension versus targeted information-seeking, and in some cases, reconstructing specific questions or goals driving the reading behavior. Shubi et al. (2025)) found that ensemble transformer-based models combining scanpath representations with LMs can predict reading goals in real-time, with performance modulated by textual properties and reader characteristics. Furthermore, Zhang et al. (2025a) demonstrated that converting raw eye-movement data into visual representations (line-graph images) and encoding them with vision transformers, temporally aligned to text via reading order, achieves good performance across multiple tasks.

These applications indicated that in practical contexts, such as automated essay scoring (AES), gaze-augmented models may improve traditional AES systems, which rely solely on text-level features (grammar, vocabulary size). Similarly, for second-language (L2) learners, gaze-augmented models may help identify which words impede comprehension. Gaze-augmented models trained on data can personalize reading materials by highlighting or simplifying difficult vocabulary in real-time. Intelligent tutoring systems can dynamically infer when a student is struggling with comprehension versus quick skimming, adjust content difficulty in response to detected cognitive load, and provide just-in-time scaffolding. In massive open online courses (MOOCs) and digital learning environments, gaze-based analytics could offer insights into learner engagement patterns, identifying moments of confusion or disengagement that might predict dropout.

3.3. Assistive Communication

Eye tracking technology has recently been increasingly used in assistive communication systems, particularly for individuals with severe speech and physical impairments. Eye gaze-based text entry (eye typing) is one of them, where the user with motor impairments points or looks at the desired letter within the user interface, where a screen keyboard and eye-tracking device are needed (Majaranta and Räihä, 2007; Panwar et al., 2012). In such assistive systems, dwell time is often used as the gaze input. However, it is likely that an incorrect input is activated when the user simply scans the interface, namely the Midas Touch problem (Jacob, 1991). To accompany a gaze-based text entry, current LMs are substantially used to enable autocompletion, context-aware predictions, and error correction during eye typing. The intended phrases can be inferred from partial input using neural language models, which also mitigates issues such as Midas Touch problem. Cai et al. (2024) introduce an LLM-assisted gaze-based text entry that saves 57% more motor actions than traditional predictive keyboards and has been tested on users with amyotrophic lateral sclerosis.

To assist users with speech disabilities and motor impairments, gaze-enabled communication platforms are designed in a way that gaze-based selection is enabled to generate speech. These Augmentation and Alternative Communication (AAC) platforms are designed to support or replace spoken and written language for individuals with speech and communication impairments. These platforms integrate gaze interfaces with symbolic communication boards, text-to-speech synthesis, and adaptive language modeling. Such systems can learn user-specific linguistic preferences, automatically generate grammatically well-formed utterances from selected semantic units, and provide context-aware topic prediction based on conversational history (Cai et al., 2024). By leveraging NLP techniques, gaze-enabled AAC platforms move beyond letter-by-letter input toward concept-based and intent-driven communication, showing the potential and future research directions for adaptive, context-aware communication technologies for non-verbal individuals (Beck Wells, 2025).

4. Discussion and Recent Challenges

A central concept that provides the appropriate framework for bridging NLP and cognitive science is the Human-in-the-Loop (HitL) framework. Although LMs effectively mimic human communication, their internal logic remains fundamentally distinct from the unobservable mental processes of the brain. By incorporating gaze data to LMs, researchers aim to move beyond simple imitation toward models that reflect human-like attention and contextual understanding. Practical applications of this paradigm are promising for personalized reading interfaces. HitL systems can adaptively simplify or enrich text to meet a reader's specific needs—a significant breakthrough for language learners or those tackling complex technical material. Gaze data may also be useful for training LMs under low-resource scenarios. In the case of training a language model with limited data, gaze information may potentially allow better generalizations about language.

However, bridging the gap between artificial attention mechanisms and human skill remains a complex frontier, presenting significant theoretical and technical challenges that must be addressed to realize these adaptive technologies fully. Consequently, numerous research questions and technical challenges have remained unsolved recently.

A main challenge in incorporating gaze data into LMs is the lack of standardization of gaze data. The lack of standardized gaze data formats limits reproducibility and data integration. In contrast to text data, which uses common schemas such as CoNLL-U or JSON, gaze datasets do not follow a shared representation. Datasets like ZuCo, GECO, CoLAGaze, and WebQAmGaze differ in sampling rates (e.g., 60Hz–1000Hz), coordinate systems (pixels, normalized space, visual angle), and preprocessing steps (fixation detection methods, saccade thresholds, outlier filtering). This variability makes cross-dataset comparison, transfer learning, and the development of general gaze-augmented models more difficult. To accelerate the adoption of gaze-augmented NLP, the community must establish a standardized format that decouples the raw eye-tracking data from the linguistic tokens.

Another challenge is the gap between human language processing, as indirectly reflected in eye movement patterns, and machine processing of language data. This gap is observable in the methodological assumptions related to using gaze data as input to LMs. Most existing studies do not directly feed raw gaze sequences into LLM prompts. Instead, LLMs are commonly used to generate linguistic labels or representations that are later compared with eye-tracking and neural measures. For example, work combining LLM-generated relevance labels with eye-tracking and EEG data shows that words marked as more relevant receive more and longer fixations, and fixation-related features can support above-chance classification of reading-related neural states (Zhang et al., 2024a,b). Recent research also reports measurable associations between LLM-derived representations and human reading behavior. Eye-tracking datasets designed for LLM evaluation show distinct fixation patterns for preferred versus rejected model responses, along with correlations between reading measures and transformer attention signals (Lopez-Cardona et al., 2025). Probing studies further indicate that internal LLM activations and attention patterns correlate with eye-tracking indicators of reading dynamics, suggesting partial alignment between model prediction processes and human behavioral signals (Wang et al., 2024).

However, empirical comparisons between human gaze patterns and transformer attention representations report only moderate alignment, with encoder models showing stronger correlations with eye-movement data than decoder models across reading tasks (Mouratidi and Poesio, 2025), suggesting a mismatch between behavioral evidence and model architecture. In addition, computational reading models that use LLM-based predictability estimates show better fits to human eye-movement data than traditional cloze-based predictability measures (Lopes Rego et al., 2024). Graph-based text structures generated by LLMs also correspond to fixation distributions, where nodes identified as more important attract higher fixation counts during reading (Zhang et al., 2025b). Alongside these approaches, an alternative direction considers providing LLMs with serialized gaze inputs, such as sequences of fixation coordinates and durations, as structured data. However, current evidence suggests that simple concatenation of numerical gaze sequences with text prompts is unlikely to produce strong performance gains without architectural mechanisms that account for the spatiotemporal nature of eye movements. This limitation arises because gaze signals are continuous, time-dependent, and spatially structured, whereas standard LLM tokenization is primarily optimized for discrete linguistic inputs. With the rapid development of multimodal LLMs and structured data tokenization methods, future systems may more effectively integrate gaze recordings through architectures that explicitly model spatiotemporal eye-movement signals alongside textual representations. A central open question is whether LLM pattern-learning and reasoning abilities extend to spatiotemporal gaze sequences when presented as structured inputs, or whether effective use of gaze data will continue to require dedicated multimodal architectures that explicitly model eye-movement dynamics.

Beyond the technical challenges and the opportunities provided by incorporating gaze data into NLP models and applications, an important aspect is ethics and privacy concerns, a responsibility unfortunately overlooked by many scholars. Eye-tracking data is increasingly treated as a sensitive biometric signal, as patterns of pupil dynamics, gaze trajectories, and eye-movement behavior can support reliable user identification and contain detailed personal information (Kröger et al., 2020; David-John et al., 2021). Beyond identification, gaze recordings may reveal a wide range of attributes without conscious user control, including neurological and behavioral disorders, cognitive load, emotional states, and psychological traits. Eye-movement features have been examined as biomarkers in research on neurodegenerative and mental health conditions, and have been linked to measurable behavioral and psychological patterns (Przybyszewski et al., 2023; Singh and Sharma, 2024; Wang et al., 2025). Machine learning methods further increase the risk of such inferences by enabling the prediction of cognitive and behavioral characteristics from high-dimensional physiological and behavioral signals (Bhatt et al., 2023). A relevant issue is the potential risk of introducing and amplifying biases. By incorporating gaze data from "majority" (cf. the "average gazer", Section 2.2), models' tendency to adopt the majority positions may be reinforced further, leading to negative social consequences.

These concerns are becoming increasingly urgent as eye-tracking technologies are rapidly integrated into consumer and mixed-reality devices, including VR/AR headsets, smartphones, webcams, and camera-based interaction systems, allowing large-scale and often passive collection of gaze data in everyday settings (Zhu et al., 2025; Bozkir et al., 2025; Liebling and Preibusch, 2014). Prior research has shown that even natural gaze behavior in virtual environments can allow user identification, highlighting the need for privacy-preserving data architectures and controlled data access mechanisms (David-John et al., 2021). Because gaze data may reveal identity, fatigue, health-related states, and cognitive or affective processes, privacy-preserving processing pipelines are necessary, particularly as intelligent systems become capable of extracting sensitive biometric and psychological information from subtle behavioral signals (Kröger et al., 2020; Bozkir et al., 2025). As eye-free applications become more capable of predicting these signals from text interaction alone, researchers must develop privacy-preserving architectures that obscure sensitive biometric signatures while retaining utility for NLP tasks.

5. Conclusions

This survey examined the role of gaze data in NLP as a cognitive signal that can support more human-aligned language modeling, tracing the development from small experimental datasets to multilingual corpora and recent synthetic gaze generation approaches. In response to the long-standing data bottleneck caused by the cost and technical constraints of eye-tracking collection, the field has gradually moved toward scalable multimodal resources and gaze synthesis methods that enable broader model training. To structure this progression, we proposed a roadmap consisting of three streams of research: constructing cognitive multimodal corpora, enriching these corpora through synthetic gaze, and training language models with gaze-informed guidance. In addition, recent approaches that incorporate gaze signals into training objectives and alignment frameworks indicate that cognitive supervision can guide model representations toward patterns observed in human reading.

Several directions follow directly from this roadmap. The reviewed studies also show practical applications in readability assessment, educational analytics, and assistive communication, where gaze-informed models can support adaptive and cognitively grounded language technologies. First, standardized evaluation protocols for synthetic gaze are needed to verify whether generated signals preserve linguistic sensitivity and reading-related patterns rather than only surface-level statistical similarity. Second, methods that retain the functional benefits of gaze data while reducing exposure of sensitive biometric patterns should be further developed, given the personal nature of cognitive signals. Third, further expansion and standardization of multilingual eye-tracking corpora is necessary, as recent resources already include datasets in languages such as Danish, German, French, and Romanian, yet cross-linguistic comparability and shared annotation standards remain limited. Finally, future work should examine whether large language models can effectively incorporate spatiotemporal gaze signals and whether dedicated multimodal architectures that explicitly model eye-movement dynamics provide more reliable performance than text-only integration strategies.

Integrating gaze data into NLP supports cognitively grounded language modeling by linking textual representations with observable reading behavior. Across corpora development, gaze synthesis, and gaze-guided training, the reviewed studies show that eye-movement signals provide measurable indicators of processing effort, attention allocation, and reading dynamics that are not captured by text-only models. While current results demonstrate the feasibility of gaze-informed modeling, future progress depends on larger multilingual corpora, standardized evaluation protocols for synthetic gaze, and architectures designed to process spatiotemporal cognitive signals in a scalable and privacy-aware manner.

6. Bibliographical References

Cengiz Acartürk. 2025. *Eyes on Text: Eye movements in reading and language processing*. John Benjamins.

Cengiz Acartürk, Jamal Nasir, Burcu Can, and Çağrı Çöltekin, editors. 2025. *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*. INCOMA Ltd., Shoumen, BULGARIA, Varna, Bulgaria.

Özge Alacam, Sanne Hoeken, and Sina Zarrieß. 2024. Eyes don't lie: Subjective hate annotation and detection with gaze. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 187–205, Miami, Florida, USA. Association for Computational Linguistics.

Diego Alves. 2025. Benchmarking language model surprisal for eye-tracking predictions in Brazilian Portuguese. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 7–17, Varna, Bulgaria. INCOMA Ltd., Shoumen, BULGARIA.

Ibtehal Baazeem, Hend Al-Khalifa, and Abdulmalik Al-Salman. 2025. Integrating linguistic and eye movements features for arabic text readability assessment using ml and dl models. *Computation*, 13(11):258.

Melissa Beck Wells. 2025. Empowering non-verbal individuals through AI-driven symbolic text prediction: a metaliteracy approach to communication and inclusion. *Discover Education*, 4(1):360.

Priya Bhatt, Amanrose Sethi, Vaibhav Tasgaonkar, Jugal Shroff, Isha Pendharkar, Aditya Desai, Pratyush Sinha, Aditya Deshpande, Gargi Joshi, Anil Rahate, et al. 2023. Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions. *Brain informatics*, 10(1):18.

Anna Bondar, David Robert Reich, and Lena Ann Jäger. 2025. CoLAGaze: A corpus of eye movements for linguistic acceptability. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, pages 1–9.

Efe Bozkir, Babette Bühler, Xiaoyuan Wu, Enkelejda Kasneci, Lujo Bauer, and Lorrie Faith Cranor. 2025. The impact of device type, data practices, and use case scenarios on privacy concerns about eye-tracked augmented reality in the United States and Germany. *Journal of Cybersecurity*, 11(1):tyaf036.

Shanqing Cai, Subhashini Venugopalan, Katie Seaver, Xiang Xiao, Katrin Tomanek, Sri Jalasutram, Meredith Ringel Morris, Shaun Kane, Ajit Narayanan, Robert L MacDonald, et al. 2024. Using large language models to accelerate communication for eye gaze typing users with als. *Nature Communications*, 15(1):9449.

Uschi Cop, Nicolas Dirix, Denis Drieghe, and Wouter Duyck. 2017. Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, 49(2):602–615.

Brendan David-John, Diane Hosfelt, Kevin Butler, and Eakta Jain. 2021. A privacy-preserving approach to streaming eye-tracking data. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2555–2565.

Shuwen Deng, David R Reich, Paul Prasse, Patrick Haller, Tobias Scheffer, and Lena A Jäger. 2023. Eyettention: An attention-based dual-sequence model for predicting human scanpaths during reading. *Proceedings of the ACM on Human-Computer Interaction*, 7(ETRA):1–24.

Ralf Engbert, André Longtin, and Reinhold Kliegl. 2002. A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, 42(5):621–636.

Ana Valeria González-Garduño and Anders Søgaard. 2017. Using gaze to predict text readability. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 438–443, Copenhagen, Denmark. Association for Computational Linguistics.

Cfir Avraham Hadar, Omer Shubi, Yoav Meiri, Amit Heshes, and Yevgeni Berzak. 2025. Decoding open-ended information seeking goals from eye movements in reading. *arXiv preprint arXiv:2505.02872*.

Anamaria Hodivoianu, Oleksandra Kuvshynova, Filip Popovici, Adrian Luca, and Sergiu Nisioi. 2025. Predicting total reading time using Romanian eye-tracking data. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 71–75, Varna, Bulgaria. INCOMA Ltd., Shoumen, BULGARIA.

Nora Hollenstein, Maria Barrett, and Lisa Beinborn. 2020a. Towards best practices for leveraging human language processing signals for natural language processing. In *Proceedings of the Second Workshop on Linguistic and Neurocognitive Resources*, pages 15–27, Marseille, France. European Language Resources Association.

Nora Hollenstein, Maria Barrett, and Marina Björnsdóttir. 2022. The copenhagen corpus of eye tracking recordings from natural reading of Danish texts. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1712–1720, Marseille, France. European Language Resources Association.

Nora Hollenstein, Marius Troendle, Ce Zhang, and Nicolas Langer. 2020b. ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 138–146, Marseille, France. European Language Resources Association.

Nora Hollenstein and Ce Zhang. 2019. Entity recognition at first sight: Improving NER with eye movement information. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1–10, Minneapolis, Minnesota. Association for Computational Linguistics.

Stephen Hutt, Aaron Wong, Alexandra Papoutsaki, Ryan S Baker, Joshua I Gold, and Caitlin Mills. 2024. Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*, 56(1):1–17.

Fariz Ikhwantri, Jan Wira Gotama Putra, Hiroaki Yamada, and Takenobu Tokunaga. 2023. Looking deep in the eyes: Investigating interpretation methods for neural models on reading tasks using human eye-movement behaviour. *Information Processing & Management*, 60(2):103195.

Oksana Ivchenko and Natalia Grabar. 2025. A French eye-tracking corpus of original and simplified medical, clinical, and general texts - FETA. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 37–43, Varna, Bulgaria. INCOMA Ltd., Shoumen, BULGARIA.

Robert JK Jacob. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)*, 9(2):152–169.

Deborah N Jakobi, Thomas Kern, David R Reich, Patrick Haller, and Lena A Jäger. 2025a. PoTeC: A German naturalistic eye-tracking-while-reading corpus. *Behavior Research Methods*, 57(8):211.

Deborah Noemie Jakobi, Maja Stegenwallner-Schütz, Nora Hollenstein, Cui Ding, Ramune Kaspere, Ana Matić Škorić, Eva Pavlinusic Vilus, Stefan Frank, Marie-Luise Müller, Kristine M Jensen de López, et al. 2025b. MultiplEYE: Creating a multilingual eye-tracking-while-reading corpus. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, pages 1–11.

Nathalie John, Sebastian P Korinth, Mareike Kunter, and Franziska Baier-Mosch. 2025. Gaze cluster analysis reveals heterogeneity in attention allocation and predicts learning outcomes. *Scientific Reports*, 15(1):20291.

Varun Khurana, Yaman Kumar, Nora Hollenstein, Rajesh Kumar, and Balaji Krishnamurthy. 2023. Synthesizing human gaze feedback for improved NLP performance. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1895–1908, Dubrovnik, Croatia. Association for Computational Linguistics.

Inbal Kimchi and Noam Siegelman. 2026. All together now: Random forests analysis reveals the joint impact of multiple statistical regularities on eye-movements during reading. *Psychonomic Bulletin & Review*, 33(1):17.

Jacob Leon Kröger, Otto Hans-Martin Lutz, and Florian Müller. 2020. What does your gaze reveal about you? on the privacy implications of eye tracking. In *IFIP International Summer School on Privacy and Identity Management*, pages 226–241. Springer.

Daniel J Liebling and Sören Preibusch. 2014. Privacy considerations for a pervasive eye tracking world. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1169–1177.

Adrielli Tina Lopes Rego, Joshua Snell, and Martijn Meeter. 2024. Language models outperform cloze predictability in a cognitive model of reading. *PLOS Computational Biology*, 20(9):e1012117.

Angela Lopez-Cardona, Sebastian Idesis, Miguel Barreda-Ángeles, Sergi Abadal, and Ioannis Arapakis. 2025. OASST-ETC dataset: alignment signals from eye-tracking analysis of LLM responses. *Proceedings of the ACM on Human-Computer Interaction*, 9(3):1–29.

Päivi Majaranta and Kari-Jouko Räihä. 2007. Text entry by gaze: Utilizing eye-tracking. In I. Scott MacKenzie and Kumiko Tanaka-Ishii, editors, *Text entry systems: Mobility, accessibility, universality*, 2007, pages 175–187. San Francisco: Morgan Kaufmann.

Danny Merkx and Stefan L. Frank. 2021. Human sentence processing: Recurrence or attention? In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 12–22, Online. Association for Computational Linguistics.

Abhijit Mishra, Kuntal Dey, and Pushpak Bhattacharyya. 2017. Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 377–387, Vancouver, Canada. Association for Computational Linguistics.

Maria Mouratidi and Massimo Poesio. 2025. Comparing eye-gaze and transformer attention mechanisms in reading tasks. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 26–36, Varna, Bulgaria. INCOMA Ltd., Shoumen, BULGARIA.

Prateek Panwar, Sayan Sarcar, and Debasis Samanta. 2012. EyeBoard: A fast and accurate eye gaze-based text entry system. In *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, pages 1–8. IEEE.

Michael I Posner, Charles R Snyder, and Brian J Davidson. 1980. Attention and the detection of signals. *Journal of experimental psychology: General*, 109(2):160.

Andrzej W Przybyszewski, Albert Śledzianowski, Artur Chudzik, Stanisław Szlufik, and Dariusz Koziorowski. 2023. Machine learning and eye movements give insights into neurodegenerative disease mechanisms. *Sensors*, 23(4):2145.

Adrielli Tina Lopes Rego, Joshua Snell, and Martijn Meeter. 2025. What determines where readers fixate next? leveraging NLP to investigate human cognition. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 1–6, Varna, Bulgaria. INCOMA Ltd., Shoumen, BULGARIA.

Erik D Reichle, Alexander Pollatsek, Donald L Fisher, and Keith Rayner. 1998. Toward a model of eye movement control in reading. *Psychological review*, 105(1):125.

Ronan G Reilly and Ralph Radach. 2003. Foundations of an interactive activation model of eye movement control in reading. In *The Mind's Eye*, pages 429–455. Elsevier.

Tiago Ribeiro, Stephanie Brandl, Anders Søgaard, and Nora Hollenstein. 2023. WebQAmGaze: A multilingual webcam eye-tracking-while-reading dataset. *arXiv preprint arXiv:2303.17876*.

Kshitij Sharma, Michail Giannakos, and Pierre Dillenbourg. 2020. Eye-tracking and artificial intelligence to enhance motivation and learning. *Smart Learning Environments*, 7(1):13.

Omer Shubi, Cfir Avraham Hadar, and Yevgeni Berzak. 2025. Decoding reading goals from eye movements. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5616–5637, Vienna, Austria. Association for Computational Linguistics.

Jaiteg Singh and Deepika Sharma. 2024. Automated detection of mental disorders using physiological signals and machine learning: A systematic review and scientometric analysis. *Multimedia Tools and Applications*, 83(29):73329–73361.

Joshua Snell, Sam van Leipsig, Jonathan Grainger, and Martijn Meeter. 2018. OB1-reader: A model of word recognition and eye movements in text reading. *Psychological review*, 125(6):969.

Ekta Sood, Prajit Dhar, Enrica Troiano, Rosy Southwell, and Sidney K. D'Mello. 2025. ScanEZ: Integrating cognitive models with self-supervised learning for spatiotemporal scanpath prediction. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1132–1142, Vienna, Austria. Association for Computational Linguistics.

Ekta Sood, Simon Tannert, Diego Frassinelli, Andreas Bulling, and Ngoc Thang Vu. 2020a. Interpreting attention models with human visual attention in machine reading comprehension. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 12–25, Online. Association for Computational Linguistics.

Ekta Sood, Simon Tannert, Philipp Müller, and Andreas Bulling. 2020b. Improving natural language processing tasks with human gaze-guided neural attention. *Advances in Neural Information Processing Systems*, 33:6327–6341.

Milan Turčáni, Zoltan Balogh, and Michal Kohútek. 2024. Evaluating computer science students reading comprehension of educational multimedia-enhanced text using scalable eye-tracking methodology. *Smart Learning Environments*, 11(1):29.

Bingbing Wang, Bin Liang, Lanjun Zhou, and Ruifeng Xu. 2024. Gaze-infused BERT: Do human gaze signals help pre-trained language models? *Neural Computing and Applications*, 36(20):12461–12482.

Min Wang, Ao Xu, Chenxiao Fan, and Xiao Sun. 2025. Machine learning for predicting personality and psychological symptoms from behavioral dynamics. *Electronics*, 14(3):583.

Francesca Zermiani, Prajit Dhar, Ekta Sood, Fabian Kögel, Andreas Bulling, and Maria Wirzberger. 2024. InteRead: An eye tracking dataset of interrupted reading. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9154–9169, Torino, Italia. ELRA and ICCL.

Dongsen Zhang, Peipei Li, Zekun Li, Yiwei Ru, Huijia Wu, and Zhaofeng He. 2025a. Eye movements as images: A multimodal framework for eye movements representation. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Yuhong Zhang, Jialu Li, Shilai Yang, Yuchen Xu, Gert Cauwenberghs, and Tzyy-Ping Jung. 2025b. Graph representations for reading comprehension analysis using large language model and eye-tracking biomarker. In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–7. IEEE.

Yuhong Zhang, Qin Li, Sujal Nahata, Tasnia Jamal, Shih-Kuen Cheng, Gert Cauwenberghs, and Tzyy-Ping Jung. 2024a. Integrating large language model, EEG, and eye-tracking for word-level neural state classification in reading comprehension. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:3465–3475.

Yuhong Zhang, Shilai Yang, Gert Cauwenberghs, and Tzyy-Ping Jung. 2024b. From word embedding to reading embedding using large language model, EEG and eye-tracking. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–4. IEEE.

Gancheng Zhu, Zehao Huang, Xiaoting Duan, Shuai Zhang, Rong Wang, Yongkai Li, and Zhiguo Wang. 2025. Smartphone eye-tracking with deep learning: Data quality and field testing. *Behavior Research Methods*, 57(7):202.