AlEYEgnment: Leveraging Eye-Tracking-while-Reading to Align Language Models with Human Preferences

Anna Bondar
Department of Computational Linguistics, Digital Society Initiative
University of Zurich
Zurich, Switzerland
anna.bondar@uzh.ch

David Robert Reich
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
davidrobert.reich@uzh.ch

Lena A. Jäger
Department of Computational Linguistics, Department of Informatics
University of Zurich
Zurich, Switzerland
lenaann.jaeger@uzh.ch

Abstract

Direct Preference Optimisation (DPO) has emerged as an effective approach for aligning large language models (LLMs) with human preferences. However, its reliance on binary feedback restricts its ability to capture nuanced human judgements. To address this limitation, we introduce a gaze-informed extension that incorporates implicit, fine-grained signals from eye-tracking-while-reading into the DPO framework. Eye movements, reflecting real-time human cognitive processing, provide fine-grained signals about the linguistic characteristics of the text that is being read. We leverage these signals and modify DPO by introducing a gaze-based additional loss term, that quantifies the differences between the model's internal sentence representations and cognitive (i.e., gaze-based) representations derived from the readers' gaze patterns. We explore the use of both human and synthetic gaze signals, employing a generative model of eye movements in reading to generate supplementary training data, ensuring the scalability of our approach. We apply the proposed approach to modelling linguistic acceptability. Experiments conducted on the CoLA dataset demonstrate performance gains in grammatical acceptability classification tasks when the models are trained in the gaze-augmented setting. These results demonstrate the utility of leveraging gaze data to align language models with human preferences. All code and data are available from Github.

1 Introduction

Direct Preference Optimisation (DPO, Rafailov et al., 2023) has recently emerged as a scalable, computationally efficient, stable method for aligning language models with human preferences. Unlike Reinforcement Learning from Human Feedback (RLHF, Christiano et al., 2017) or Reinforcement Learning from AI Feedback (RLAIF, Lee et al., 2024), it does not require a separate reward model and allows a policy model to internalise the preferences directly. However, DPO relies only on binary pairs of preferred and dispreferred responses, and this simplicity leads to a critical limitation: binary feedback provides no information about how strongly one response is preferred over another, limiting the model's ability to align with nuanced human judgements. Recent studies have demonstrated that integrating explicit, fine-grained preference labels—such as ranked lists or ordinal scores— into a DPO-based framework improves the alignment of a policy model with human preferences (Liu et al., 2025; Zhao et al., 2024). However, collecting explicit high-quality detailed annotations from humans at scale is labour-intensive and costly.

To address the outlined limitations, we introduce a method that leverages implicit human feedback from eye-tracking data collected during reading. Eye movements are considered the gold-standard method to investigate cognitive processes underlying language processing (Rayner, 1998; Clifton et al., 2007). Because eye movement patterns systematically reflect processing difficulty and readers' evaluations of linguistic input, these gaze signals can provide detailed, fine-grained indicators of human preferences, reducing reliance on explicit detailed human ranking or rewards from auxiliary models. In our approach, we integrate gaze-based signals into the DPO training pipeline, allowing the model to incorporate nuanced human feedback beyond binary supervision while retaining DPO's computational efficiency. Recent advances in generative models of eye movements in reading further support the scalability of this method (Prasse et al., 2023; Deng et al., 2023b; Bolliger et al., 2023, 2025) since they make it possible to generate synthetic human-like scanpaths and increase the training dataset without collecting data from humans.

The specific downstream task we focus on is

Proceedings of First International Workshop on Gaze Data and NLP associated with RANLP 2025, pages 58–70, Varna, Bulgaria, Sep 12, 2025.
https://doi.org/10.26615/978-954-452-104-2-008

modelling grammatical acceptability (terms grammaticality and acceptability are used interchangeably) judgements. Models for this task are typically trained using supervised learning with binary labels that categorise sentences into acceptable versus unacceptable ones. However, psycholinguistic research (Lau et al., 2017; Francis, 2021) demonstrates that speakers perceive grammatical acceptability along a gradient rather than as a binary label. Eye-tracking data holds potential to inform the model about the degree of ungrammaticality, as psycholinguistic evidence demonstrates that eye movement patterns vary depending on the degree of grammar violations (Tuninetti et al., 2015; Rayner et al., 2004). Similarly, eye movements in reading can provide information on the type, strength of ungrammaticality (Braze et al., 2002) and its location within the utterance (Vasishth et al., 2013; Frazier and Rayner, 1982). Once an ungrammaticality is encountered, the reader's eye-movement patterns tend to exhibit longer fixation durations, an increased number of regressions, and disrupted saccadic movements—reflecting increased processing difficulty and reanalysis effort. These findings suggest that eye-tracking data can supply the fine-grained online signal missing from binary annotations and inform the model about the strength, locus, and characteristics of grammar violations as perceived by humans. Building on this foundation, we investigate two principal research questions: (i) whether integrating gaze signals into DPO during training improves model performance on grammatical acceptability; and (ii) whether increasing the amount of training data by adding synthetic gaze data leads to further gains in model performance.

2 Related Work

2.1 Human-Preferences Alignment

Direct Preference Optimisation has been proposed as a streamlined alternative to RLHF and RLAIF, as it trains directly on binary preferred–dispreferred pairs and does not require a learned reward model (Rafailov et al., 2023). However, this pairwise supervision limits the model's capacity to reflect how strongly one response is preferred over another. Recent work has introduced methods to incorporate finer-grained information. One of these methods—Ordinal Preference Optimisation (OPO)—replaces binary comparisons with ranked lists, enabling the model to capture relative distances among responses (Zhao et al., 2024). Another approach—Listwise Preference Optimisation (LiPO)—extends this idea by formulating alignment as a learning-to-rank problem (Liu et al., 2025). An alternative method, namely Margin Matching Preference Optimisation (MMPO), retains the pairwise format of the responses and attaches real-valued quality margins to each pair (Kim et al., 2024). All of these approaches rely on explicit, graded feedback, either from human annotations or reward models, which can be costly to obtain or may diverge from human judgements when using external LLMs to score the responses (Bavaresco et al., 2025).

2.2 Eye Movements in Reading as Indicators of Grammatical Violations

Eye-tracking studies have demonstrated that gaze patterns reliably identify the locus, type, and strength of grammatical violations (Schotter and Dillon, 2025). Readers precisely localise syntactic anomalies, leading to immediate regressions and increased fixation durations at points of structural disambiguation or grammatical inconsistency (Frazier and Rayner, 1982; Vasishth et al., 2013). Furthermore, distinct gaze signatures differentiate violation types: syntactic errors (e.g., agreement mismatches or structural ambiguities) typically cause rapid regressions and increased first fixation durations, whereas semantic and pragmatic anomalies predominantly affect later reading measures, such as regression-path duration and total fixation time (Braze et al., 2002).

Eye movements also systematically reflect the strength of violation. Strong violations, such as outright ungrammatical constructions or semantically impossible continuations, provoke immediate disruptions in first-pass reading times and frequent regressions. Conversely, milder violations, such as subtle semantic implausibilities or pragmatic errors, result in delayed and comparatively moderate reading disruptions, evident primarily through increased regression-path durations and cumulative reading times (Rayner et al., 2004; Tuninetti et al., 2015; Joseph et al., 2009; Schotter and Dillon, 2025; Schotter and Jia, 2016). Overall, these findings demonstrate the utility of eye-tracking as a fine-grained implicit feedback on the processing of grammatical violations in real-time language comprehension.

2.3 Eye-Tracking-while-Reading for Natural Language Processing

Eye movements in reading have been leveraged for model evaluation and interpretation, including the assessment of a model's and cognitive plausibility (Bolliger et al., 2024; Beinborn and Hollenstein, 2024; Haller et al., 2024; Goodkind and Bicknell, 2018; Bensemann et al., 2022; Eberle et al., 2022; Sood et al., 2020a; Hollenstein and Beinborn, 2021).

Besides model evaluation and interpretation, gaze signals have proven effective for training and evaluating NLP models. Recent research demonstrated that eye movements in reading can be leveraged as a supervisory signal to enhance model performance on various downstream NLP tasks. Early research employed eye-tracking data in the form of auxiliary input alongside the text embeddings for named entity recognition, sentiment analysis, sarcasm detection, part-of-speech tagging (Hollenstein and Zhang, 2019; Mishra et al., 2016; Barrett et al., 2016; Tiwari et al., 2023). Other studies integrated reading measures into models to guide attention mechanism directly for visual question answering, sentence compression and paraphrase generation, sentiment analysis (Sood et al., 2020b; Long et al., 2017; Sood et al., 2023). Further research utilised gaze data in transfer learning settings, tasking the models to predict reading measures as an auxiliary training objective for sarcasm detection, readability prediction, or machine reading comprehension (Yang and Hollenstein, 2023; Deng et al., 2023a; González-Garduño and Søgaard, 2018; Malmaud et al., 2020). A more recent line of research reordered the input sequence according to the scanpaths (Yang and Hollenstein, 2023; Deng et al., 2024) at the fine-tuning stage. All of the listed frameworks have demonstrated the utility of eye movements in reading for a wide range of NLP tasks and have exhibited comparable performance using either real human or synthetic eye-tracking data.

Most recently, eye-tracking data has been integrated into frameworks aimed at aligning human preferences, specifically in reward modelling within RLHF paradigms (López-Cardona et al., 2025). Eye movements have also shown promising results for constructing datasets reflecting human preferences (Kiegeland et al., 2024; Lopez-Cardona et al., 2025). Nevertheless, directly applying gaze data to preference alignment frameworks without relying on intermediate reward models remains unexplored. We address this gap and demonstrate the utility of eye-tracking-while-reading data for directly aligning large language models with human preferences.

3 Preliminaries

We first provide a short overview of Direct Preference Optimisation (DPO, Rafailov et al., 2023), a method for aligning language models with human preferences. This approach is a further development of RLHF (Ouyang et al., 2022) and relies on a policy model—the model being trained—a reference model—a frozen, pre-trained checkpoint used to regularise training and keep the policy close to its initial weights—and a reward model, which assigns rewards to outputs produced by the policy. DPO eliminates this reward model and instead optimises the policy to increase the (log-)probability of preferred over dispreferred responses directly.

Given a dataset of triples (r, x^1, x^0)—a prompt r with a preferred (chosen) response x^1 and a dispreferred (rejected) response x^0—DPO updates a policy π_θ relative to a fixed reference policy π_{ref} by maximising the Bradley–Terry log-likelihood:

$$\max_\theta \; \mathbb{E}_{(r,x^1,x^0)}\Big[\log\sigma\Big(\beta\Big(\log\frac{\pi_\theta(x^1|r)}{\pi_{\text{ref}}(x^1|r)} - \log\frac{\pi_\theta(x^0|r)}{\pi_{\text{ref}}(x^0|r)}\Big)\Big)\Big], \tag{1}$$

where σ is a sigmoid function that maps the difference in log-probability ratios between the policy and reference models to a value in $(0, 1)$, which can be interpreted as the probability that the policy assigns a higher probability to the preferred response x^1, β is a temperature parameter that controls the sensitivity of the model to small differences between the preferred and dispreferred options. This objective directly increases the model's relative log-probability of preferred over dispreferred responses.

4 Problem Setting

The task of linguistic acceptability classification is a supervised learning problem, where the goal is to determine whether a given natural language expression conforms to the grammatical norms of a particular language variety. Formally, let $\mathcal{X} \subset \Sigma^*$ denote all possible input strings over a finite vocabulary Σ. Each input $x \in \mathcal{X}$ is a sentence. The output is $\mathcal{Y} \in \{0, 1\}$, where $y = 1$ indicates an acceptable expression and $y = 0$ denotes an unacceptable one. Given a dataset $\mathcal{D} = \{(x_n, y_n)\}_{n=1}^{N}$ sampled from $\mathcal{X} \times \mathcal{Y}$, the objective is to find a function $f_\theta : \mathcal{X} \to [0, 1]$ parametrised by θ, where $f_\theta(x)$ represents the predicted probability of acceptability. We investigate two questions: (i) whether incorporating human eye-tracking signals at training time improves performance on grammatical acceptability classification, and (ii) whether adding synthetic gaze provides further gains beyond human signals alone.

We evaluate our models and report the performance with accuracy, F_1, and Matthews correlation coefficient on held-out data.

5 Data

5.1 The CoLAGaze Corpus

We utilised the CoLAGaze eye-tracking-while-reading corpus (Bondar et al., 2025) to integrate implicit human feedback into the Direct Preference Optimisation framework. The dataset comprises eye-tracking data collected from 42 participants reading 153 pairs of (un)grammatical sentences manually selected from the Corpus of Linguistic Acceptability (CoLA; Warstadt et al., 2019). Each participant read either the grammatical or the ungrammatical counterpart of each sentence. These sentences span a diverse range of grammatical violations, including syntactic, morphosyntactic, semantic, and pragmatic anomalies. Detailed information on the original data collection procedure, preprocessing steps, and computation of reading measures can be found in Bondar et al. (2025). The full corpus contains 6,246 data points in total. For our analyses, we selected data from 38 well-calibrated participants, resulting in a total of 5,814 data points.

5.2 Synthetic Data

In addition to the human data provided by CoLAGaze, we trained Eyettention (Deng et al., 2023b), a state-of-the-art generative model of eye movements in reading, to produce synthetic scanpaths (i.e., sequences of fixations and saccades) for an additional 30 sentence pairs from CoLA (see Appendix B for more details), with the goal of extending the training set beyond the original CoLAGaze data and assess whether gaze-informed models can benefit from synthetic gaze signals during training.

6 Method

6.1 DPO with Binary Feedback

To address the linguistic acceptability classification task, we fine-tune a 7-billion-parameter instruction-tuned Mistral model within the Direct Preference Optimisation framework. To fit the DPO setup we form a set of 153 pairwise preferences $\mathcal{P} = \left\{\left(x_c^1, x_c^0\right)\right\}_{c=1}^{C}$, where x_c^1 represents a grammatical sentence and x_c^0 its ungrammatical counterpart, and C represents the total number of pairs. The DPO setup employs two pretrained LLMs: the policy model π_θ, initialised from the instruction-tuned Mistral checkpoint and fine-tuned during training, and the reference model π_{ref}, which shares the initial parameters of the policy model but remains frozen throughout fine-tuning to stabilise learning and avoid catastrophic forgetting. Given a prompt r that explicitly instructs the model to identify the grammatical sentence from the pair (x_c^1, x_c^0) (for details on the prompt, see Section 7), the policy model is trained to generate the grammatical sentence as output. We optimise the parameters of the model using the standard DPO objective (for details on standard DPO, see Section 3).

6.2 DPO Augmented with Gaze-Based Implicit Feedback

6.2.1 Eye-Tracking Feature Selection

We used sentence-level eye-tracking measures from CoLAGaze calculated after correcting for vertical drift. Specifically, we selected a subset of eye-tracking features most predictive of sentence-level acceptability across all violation types included in CoLAGaze. To select the subset from the CoLAGaze dataset, we fit a binomial generalised linear mixed model to predict sentence labels from the eye-tracking features and perform greedy backward (recursive) elimination, removing one feature at a time and refitting the model. Feature selection is guided by the Bayesian Information Criterion (BIC) (Schwarz, 1978). The final set of features is the one that minimises BIC.

Once the subset of the eye-tracking features is selected, we train our gaze-augmented large language models with two sets of eye-tracking measures (see Appendix C for a comprehensive list of measures and their definitions): measures based on event counts (e.g. number of fixations, number of regressions) and measures based on durations (e.g. total fixation duration, first-pass reading time). Models augmented with synthetic eye-tracking data utilise only event-count based features, as the Eyettention model employed for synthetic data generation does not predict fixation durations.

6.2.2 Integration of the Eye-Tracking Data

To integrate the cognitive information into the DPO framework, we introduce an additional gaze-based loss term ℓ_{ET} to the original DPO loss function, that quantifies the alignment of the model's internal sentence representations h to cognitive (i.e., gaze-based) representations g derived from the sequence of eye-movement events s (see Figure 1 for a visualisation of how the gaze-based loss term is derived). To compute the eye-tracking based loss term, for a grammatical–ungrammatical pair (x_c^1, x_c^0), we obtain the sentence embeddings $h_c^1, h_c^0 \in \mathbb{R}^d$ from the policy model π_θ. To get the embeddings we tokenise the two sentences from each pair into two separate sequences $T_{x_c} = \{\, t_1, \ldots, t_{|T_{x_c}|} \,\}$, feed each of the sequences to the model π_θ, extract the hidden states of the last layer $H \in \mathbb{R}^{T_x \times d}$ from the model and use mean pooling to derive a sentence representation

$$h_c \;=\; \frac{1}{T_{x_c}} \sum_{q=1}^{T_{x_c}} H_q, \tag{2}$$

where q is a token position in a sequence. To integrate gaze data into the loss, we form eye-tracking feature vectors g_c, consisting of the selected sentence-level eye-tracking features from CoLAGaze (see 6.2.1). Let I_c^1 and I_c^0 denote the set of readers who saw x_c^1 and x_c^0, respectively[1]; for each reader $i \in I_c^1$, or $j \in I_c^0$, and for each sentence x_c^1, or x_c^0, we form a sentence-level gaze feature vector $g_{c,i}^1 \in \mathbb{R}^F$, or $g_{c,j}^0 \in \mathbb{R}^F$, where F is the number of gaze features. For each sentence pair (x_c^1, x_c^0) we form $K = 20$ cross-participant vector pairs by independently sampling indices $i_k \in I_c^1$ and $j_k \in I_c^0$ with replacement for each pair; these indices are fixed once at the start of training. (we treat the number of pairs K as a hyperparameter, see Appendix A for details) and compute the difference between them $\Delta_{\text{gaze}_c}^{(k)} = g_{c,i_k}^1 - g_{c,j_k}^0$. We then treat each tuple consisting of the prompt, the grammatical and ungrammatical sentences and the gaze vector difference $(r, x_c^1, x_c^0, \Delta_{\text{gaze}_c}^{(k)})$ as a separate gaze-augmented training instance. For each of the instances we compute the gaze-based loss term

$$\ell_{\text{ET}_c}^{(k)} = \cos(h_c^1, h_c^0)\; \big\|\Delta_{\text{gaze}_c}^{(k)}\big\|_2. \tag{3}$$

Because $\cos(h_c^1, h_c^0)$ is in the range $[-1, 1]$, ℓ_{ET_c} penalises the model when the sentence embeddings are too similar while the differences in human gaze patterns are large — in this case the cosine term is close to 1 and the Euclidean distance between the gaze vectors $\big\|\Delta_{\text{gaze}_c}^{(k)}\big\|_2$ is large, this results into large positive gaze-based loss. On the other hand, when the sentence representations are already well separated (cosine term closer to -1), the gaze-based loss term $\ell_{\text{ET}_c}^{(k)}$ becomes negative and implicitly rewards the model by decreasing the total loss. Training minimises the expectation over the final loss:

$$\mathcal{L}_{\text{total}} = \mathbb{E}_{(r,x^1,x^0,\Delta_{\text{gaze}})}\big[\, \mathcal{L}_{\text{DPO}}(\theta) + \alpha\, \ell_{\text{ET}} \,\big], \tag{4}$$

where α is a tuned hyperparameter. By training the model with a gaze-based loss term we intend to align the model's representations with human cognitive processing signals.

7 Experiments

7.1 Training Setup

We fine-tuned the 7-billion-parameter instruction-tuned Mistral model in several configurations to evaluate whether integrating implicit feedback derived from eye-tracking data into Direct Preference Optimisation enhances downstream performance on grammatical acceptability classification. See Figure 2 for a summary of the training and evaluation pipeline. Training details are available in Appendix A.

We model grammatical acceptability as a binary classification task, implemented as text generation with a decoder-only transformer. At training, for each item, both grammatical and ungrammatical sentences are presented in a single prompt:

```
Select the grammatically correct
sentence:
A) <sent_A>
B) <sent_B>
```

The assignment of the grammatical option to A or B is random to avoid position cues. The policy model π_θ computes log-probabilities for each sentence; grammatical sentences are treated as preferred responses and ungrammatical ones as dispreferred.

We augmented the DPO framework with gaze data in several configurations. First, we incorporated implicit human gaze feedback, where the

[1]The stimuli were presented in Latin square such that each reader saw either grammatical or ungrammatical version of each sentence

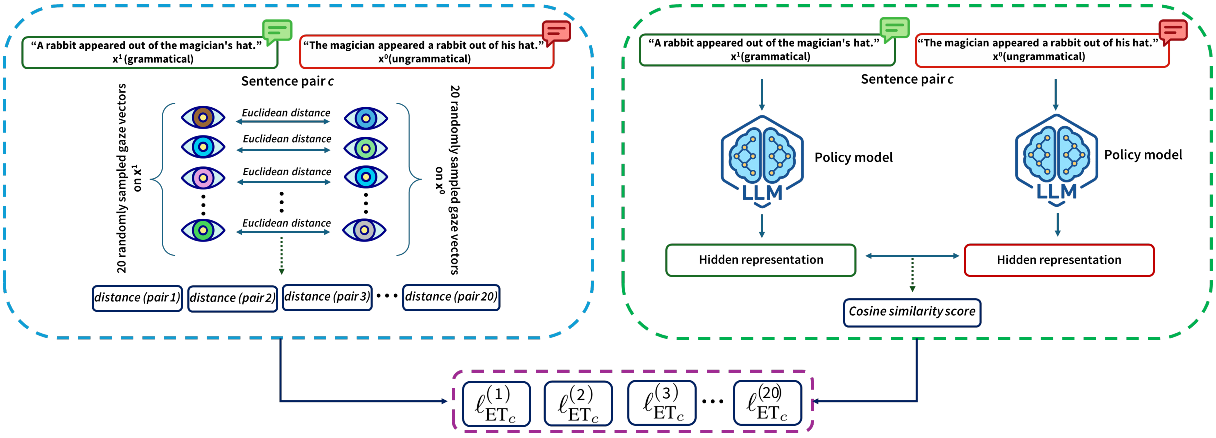

Figure 1: As depicted in the left blue box, for each grammatical–ungrammatical sentence pair c, we randomly sample 20 gaze vectors per grammatical sentence x^1 and per ungrammatical one x^0, drawn from different readers. From the sampled gaze vectors we then randomly pick a gaze vector for the grammatical sentence and a gaze vector for its ungrammatical counterpart and pair these gaze vectors, resulting into 20 gaze vector pairs. For each pair of gaze vectors, we compute the Euclidean distance between them to quantify the difference in gaze behaviour for the grammatical sentence compared to its ungrammatical counterpart. As shown in the right green box, we simultaneously, extract hidden representations of the sentences from the policy model and calculate the cosine similarity between the grammatical and ungrammatical sentences, reflecting their proximity in the model's internal representation space. Each gaze distance is then multiplied by the similarity of the hidden representations to produce a scalar eye-tracking-based loss $\ell^{(k)}_{\mathrm{ET}_c}$.

eye-tracking-based loss was derived from selected CoLAGaze features, using both event durations and event-count measures. Second, we experimented with using only count based features to assess whether duration based measures contribute to performance gains. Third, we extended this setup by incorporating synthetic eye-tracking data, by adding synthetic scanpaths on 30 additional sentence pairs generated by the Eyettention model (see B for details). Finally, we investigated whether averaging gaze features across all readers—representing an "average reader"—still leads to improved performance.

7.2 Baselines

We evaluated our method against three text-only baselines based on the instruction-tuned Mistral checkpoint. First, the Base model corresponds to the original checkpoint without any task-specific fine-tuning.

Second, we trained a Supervised Fine-Tuning (SFT) variant by optimising a cross-entropy loss on the 153 grammatical–ungrammatical sentence pairs from CoLA. When the policy model is trained in the SFT setting, it is fine-tuned to generate the acceptability label y from a prompt t containing a sentence x_n to be classified with a labely_n as being either grammatical (1) or ungrammatical (0) and a question *"Is this sentence grammatical?"*. SFT minimises the cross-entropy loss (negative log-likelihood) over dataset $\mathcal{D}$ defined in Section 4:

$$\mathcal{L}_{\mathrm{SFT}}(\theta) = -\,\mathbb{E}_{(x,y)\sim\mathcal{D}}\big[\log \pi_\theta\big(y \mid t, x\big)\big] \approx -\,\frac{1}{N}\sum_{n=1}^{N}\log \pi_\theta\big(y \mid t, x_n\big). \tag{5}$$

Third, we trained a text-only DPO model (see Section 6.1 for details) using the same sentence pairs as in the previous training settings, relying solely on binary acceptability supervision without any cognitive signals.

7.3 Ablation

To further validate our findings, we conduct an ablation study eliminating the eye-tracking features in the additional loss term ℓ. In this variant, the standard DPO objective is augmented only with the cosine similarity between the two sentence embeddings, $\cos(h_c^1, h_c^0)$, omitting the gaze-difference term (i.e., effectively setting $\|\Delta^{(k)}_{\mathrm{gaze}_c}\|_2 = 0$).

7.4 Evaluation Setup

At test time, we use only the text data from the held-out CoLA training and development sets. Each test sentence is fed to the model alongside the following prompt:

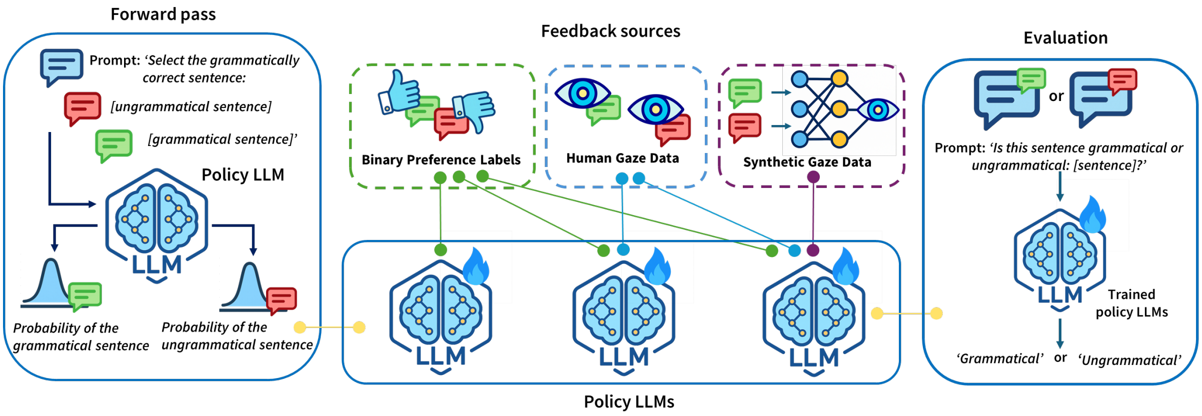

Figure 2: As depicted in the left part of the figure, during DPO training, the model receives a prompt containing a pair of sentences—one grammatical and one ungrammatical—and is tasked to select the grammatical one. We extract the probability of the grammatical and ungrammatical sentence being generated and use them for DPO. As the central block of the figure shows, we fine-tune the model and explore three training configurations: (i) using only binary preference labels (text-only), (ii) augmenting the labels with human gaze data, and (iii) further incorporating synthetic gaze data into the training. At evaluation, shown in the right side of the figure, the model is given a single sentence and prompted to classify it as grammatical or ungrammatical. Eye-tracking data is not used during evaluation.

```
Is this sentence grammatical or
ungrammatical?  <sent>
```

We report results separately on two subsets (both sourced from CoLA training and CoLA development set): sentences that share linguistic characteristics with the training data, such as similar syntactic constructions and lexical items (*in-domain* subset in the original CoLA dataset), and those that differ substantially from the training distribution (*out-of-domain* subset in the original CoLA dataset). Performance is measured using accuracy, F1 score, and Matthews correlation coefficient (MCC).

8 Results and Discussion

The results for grammatical acceptability classification on CoLA are summarised in Table 1.

Supervised Fine-Tuning performed notably worse than the instruction-tuned base model, obtaining an MCC of 0.463 in-domain and 0.410 out-of-domain. The model trained with text-only DPO also failed to surpass the base model's performance with an MCC of 0.460 and 0.406 for in-domain and out-of-domain, respectively. This drop in performance could be attributed in part to the small size of the dataset used for training (fine-tuning or DPO), which may have led the model to overfit and generalise poorly (Barnett et al., 2024). Additionally, the Supervised Fine-Tuning and DPO training was conducted using quantised low-rank adaptation (QLoRA, Dettmers et al., 2023), potentially further limiting the effective model capacity (Wang et al., 2024).

The best-performing model was Mistral fine-tuned using DPO augmented with eye-tracking features—both event-count and duration based. This model achieved an MCC of 0.510 in-domain and 0.502 out-of-domain. Relative to the baseline instruction-tuned Mistral model, gaze-augmented DPO improved the MCC by 0.037 points in-domain and 0.074 points out-of-domain. Similar improvements were observed for F1 and accuracy metrics. These results indicate benefits from integrating eye-tracking signals into the optimisation objective. In particular, this method appears useful in a low resource settings, as in our study the models were trained on only 153 sentence pairs. Finally, our ablation study demonstrates that, as expected, removing the gaze signal leads to inferior model performance.

We further compared gaze augmentation with all gaze-derived features against leveraging a reduced set containing only fixation- and saccade-count based features. The results showed an advantage of using all gaze features, suggesting that duration based gaze features contribute additional information beyond fixation counts alone.

Overall, the model trained with both gaze-event-count and duration based features outperformed the baseline and the models trained on text only. The results hold for all of the settings in which

Model	Gaze Data	Synthetic Data	Aggregated	Test Set	Accuracy↑	F1↑	MCC↑
Base	✗	✗	✗	in-domain	76.62	0.83	0.473
Base	✗	✗	✗	out-of-domain	72.87	0.79	0.428
SFT	✗	✗	✗	in-domain	$70.40_{2.45}$	$0.750_{0.033}$	$0.463_{0.009}$
SFT	✗	✗	✗	out-of-domain	$64.87_{2.36}$	$0.678_{0.037}$	$0.410_{0.019}$
DPO	✗	✗	✗	in-domain	$72.60_{0.19}$	$0.778_{0.004}$	$0.460_{0.006}$
DPO	✗	✗	✗	out-of-domain	$67.20_{0.41}$	$0.734_{0.03}$	$0.406_{0.008}$
Ablation	✗	✗	✗	in-domain	$76.24_{0.54}$	$0.8217_{0.0053}$	$0.4711_{0.0058}$
Ablation	✗	✗	✗	out-of-domain	$73.74_{0.68}$	$0.7929_{0.0083}$	$0.4521_{0.0029}$
DPO	all feat-s	✗	✗	in-domain	$80.17_{0.667}$	$0.864_{0.006}$	$0.510_{0.013}$
DPO	all feat-s	✗	✗	out-of-domain	$79.33_{0.872}$	$0.855_{0.008}$	$0.502_{0.008}$
DPO	count feat-s	✗	✗	in-domain	$79.53_{0.110}$	$0.858_{0.001}$	$0.499_{0.002}$
DPO	count feat-s	✗	✗	out-of-domain	$78.79_{0.469}$	$0.848_{0.004}$	$0.496_{0.009}$
DPO	all feat-s	✗	✓	in-domain	$79.76_{0.135}$	$0.868_{0.002}$	$0.490_{0.005}$
DPO	all feat-s	✗	✓	out-of-domain	$78.15_{0.070}$	$0.854_{0.001}$	$0.455_{0.001}$
DPO	count feat-s	✗	✓	in-domain	$76.24_{0.288}$	$0.821_{0.007}$	$0.472_{0.001}$
DPO	count feat-s	✗	✓	out-of-domain	$73.06_{0.551}$	$0.788_{0.008}$	$0.436_{0.003}$
DPO	count feat-s	✓	✓	in-domain	$76.15_{0.134}$	$0.821_{0.002}$	$0.471_{0.002}$
DPO	count feat-s	✓	✓	out-of-domain	$73.45_{0.820}$	$0.791_{0.009}$	$0.446_{0.007}$

Table 1: Results of training Mistral model on 153 sentence pairs from CoLA in different configurations: in-domain and out-of-domain subsets. Accuracy, F1 (positive = grammatical), and MCC are reported as mean_{SD} over 3 random seeds. *Gaze Data* indicates whether human eye-tracking features were used; *Synthetic Data* indicates whether synthetic gaze features were additonally used; *Aggregated* refers to whether gaze features were aggregated across readers. *All feat-s* in the *Gaze Data* columns means that both duration and event-count based features were used at training, *count feat-s* means that only event-count based features were leveraged.

the gaze data with all of the features was used — augmented with the data not aggregated across the readers, and with scanpaths aggregated across the readers. These findings are in line with the seminal work by Kliegl et al. (1982), who first showed that both duration and event-count based measures are informative about processing difficulty. The DPO training with event-count based features does not consistently lead to performance gains — while using the data not aggregated across the readers is beneficial, aggregating across participants leads to a decrease in performance in in-domain evaluation settings. Models where training was augmented with synthetic gaze data showed only marginal improvements over the base model on the out-of-domain test set. We attribute this to several factors, namely the small size of the synthetic dataset, the usage of a single gaze-feature vector per sentence, and the reliance on event-count reading measures only. Future research might investigate the integration of synthetic data with both duration and event-count based features, and explore the use of larger synthetic datasets.

Finally, future work may examine word-level eye-tracking features instead of sentence-level measures, as these have the potential to localise ungrammaticality within sentences and thereby provide the model with a more fine-grained and informative supervision signal.

9 Conclusion

We introduced a gaze-informed extension of Direct Preference Optimisation that aligns a large language model's internal representations with human cognitive processing signals. By integrating an eye-tracking loss term—derived from sentence-level differences in reading patterns observed on grammatical versus ungrammatical sentences—into the DPO objective, our approach injects graded, implicit feedback into training. Our experiments on CoLAGaze show that gaze-augmented models consistently outperform text-only baselines, and that both duration-based and count-based eye-tracking features provide useful signals beyond text alone.

Acknowledgements

This work was partially funded by the Digital Society Initiative at the University of Zurich via a PhD-scholarship granted to Anna Bondar, and the Swiss National Science Foundation under grant IZCOZ0_220330 (EyeNLG, PI: Lena Jäger). It was furthermore supported by the European Cooperation in Science and Technology (COST) under COST Action CA21131 (MultiplEYE).

References

Scott Barnett, Zac Brannelly, Stefanus Kurniawan, and Sheng Wong. 2024. Fine-Tuning or Fine-Failing?

Debunking Performance Myths in Large Language Models. *arXiv*, 2406.11201.

Maria Barrett, Joachim Bingel, Frank Keller, and Anders Søgaard. 2016. Weakly Supervised Part-of-speech Tagging Using Eye-tracking Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 579–584, Berlin, Germany.

Anna Bavaresco, Raffaella Bernardi, Leonardo Bertolazzi, Desmond Elliott, Raquel Fernández, Albert Gatt, Esam Ghaleb, Mario Giulianelli, Michael Hanna, Alexander Koller, Andre Martins, Philipp Mondorf, Vera Neplenbroek, Sandro Pezzelle, Barbara Plank, David Schlangen, Alessandro Suglia, Aditya Surikuchi, Ece Takmaz, and Alberto Testoni. 2025. LLMs instead of Human Judges? A Large Scale Empirical Study across 20 NLP Evaluation Tasks. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 238–255, Vienna, Austria.

Lisa Beinborn and Nora Hollenstein. 2024. *Cognitive Plausibility in Natural Language Processing*. Synthesis Lectures on Human Language Technologies. Springer Nature, Cham, Switzerland.

Joshua Bensemann, Alex Peng, Diana Benavides-Prado, Yang Chen, Neset Tan, Paul Michael Corballis, Patricia Riddle, and Michael Witbrock. 2022. Eye Gaze and Self-attention: How Humans and Transformers Attend Words in Sentences. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 75–87, Dublin, Ireland.

Yevgeni Berzak, Boris Katz, and Roger Levy. 2018. Assessing Language Proficiency from Eye Movements in Reading. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 1986–1996, New Orleans, Louisiana.

Yevgeni Berzak, Chie Nakamura, Suzanne Flynn, and Boris Katz. 2017. Predicting Native Language from Gaze. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 541–551, Vancouver, Canada.

Yevgeni Berzak, Chie Nakamura, Amelia Smith, Emily Weng, Boris Katz, Suzanne Flynn, and Roger Levy. 2022. CELER: A 365-Participant Corpus of Eye Movements in L1 and L2 English Reading. *Open Mind*, 6:41–50.

Lena Bolliger, Patrick Haller, and Lena Jäger. 2024. On the Alignment of LM Language Generation and Human Language Comprehension. In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pages 217–231, Miami, FL.

Lena Bolliger, David Reich, Patrick Haller, Deborah Jakobi, Paul Prasse, and Lena Jäger. 2023. ScanDL: A Diffusion Model for Generating Synthetic Scanpaths on Texts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, page 15513–15538, Singapore.

Lena Bolliger, David Reich, and Lena Jäger. 2025. ScanDL 2.0: A Generative Model of Eye Movements in Reading Synthesizing Scanpaths and Fixation Durations. In *Proceedings of the ACM on Human-Computer Interaction*, 3, New York, NY.

Anna Bondar, David Reich, and Lena Jäger. 2025. CoLAGaze: A Corpus of Eye Movements for Linguistic Acceptability. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, New York, NY. Association for Computing Machinery.

David Braze, Donald Shankweiler, Weijia Ni, and Laura Conway Palumbo. 2002. Readers' Eye Movements Distinguish Anomalies of Form and Content. *Journal of Psycholinguistic Research*, 31(1):25–44.

Paul Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep Reinforcement Learning from Human Preferences. In *Proceedings of Advances in Neural Information Processing Systems 30*, pages 4299–4311, Long Beach, CA.

Charles Clifton, Adrian Staub, and Keith Rayner. 2007. Eye Movements in Reading Words and Sentences. In Roger Van Gompel, Martin Fischer, Wayne Murray, and Robin Hill, editors, *Eye Movements*, pages 341–371. Elsevier, Netherlands.

Shuwen Deng, Paul Prasse, David Reich, Tobias Scheffer, and Lena Jäger. 2023a. Pre-trained Language Models Augmented with Synthetic Scanpaths for Natural Language Understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6500–6507, Singapore.

Shuwen Deng, Paul Prasse, David Reich, Tobias Scheffer, and Lena Jäger. 2024. Fine-tuning Pre-trained Language Models with Gaze Supervision. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pages 217–224, Bangkok, Thailand.

Shuwen Deng, David Reich, Paul Prasse, Patrick Haller, Tobias Scheffer, and Lena Jäger. 2023b. Eyettention: An Attention-based Dual-Sequence Model for Predicting Human Scanpaths during Reading. In *Proceedings of the ACM on Human-Computer Interaction*, 162, New York, NY.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient Fine-tuning of Quantized LLMs. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, New Orleans, LA.

Oliver Eberle, Stephanie Brandl, Jonas Pilot, and Anders Søgaard. 2022. Do Transformer Models Show Similar Attention Patterns to Task-Specific Human Gaze? In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4295–4309, Dublin, Ireland.

Elaine Francis. 2021. The Problem of Gradient Acceptability. In *Gradient Acceptability and Linguistic Theory*. Oxford University Press, Oxford, United Kingdom.

Lyn Frazier and Keith Rayner. 1982. Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences. *Cognitive Psychology*, 14(2):178–210.

Ana González-Garduño and Anders Søgaard. 2018. Learning to Predict Readability Using Eye-movement Data from Natives and Learners. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, New Orleans, LA.

Adam Goodkind and Klinton Bicknell. 2018. Predictive Power of Word Surprisal for Reading Times is a Linear Function of Language Model Quality. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics*, pages 10–18, Salt Lake City, UT.

Patrick Haller, Lena Bolliger, and Lena Jäger. 2024. Language Models Emulate Certain Cognitive Profiles: An Investigation of How Predictability Measures Interact with Individual Differences. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 7878–7892, Bangkok, Thailand.

Nora Hollenstein and Lisa Beinborn. 2021. Relative Importance in Sentence Processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 141–150, Online.

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, MN.

Holly Joseph, Simon Liversedge, Hazel Blythe, Sarah White, and Keith Rayner. 2009. Word Length and Landing Position Effects during Reading in Children and Adults. *Vision Research*, 49(16):2078–2086.

Samuel Kiegeland, David Reich, Ryan Cotterell, Lena Jäger, and Ethan Wilcox. 2024. The Pupil Becomes the Master: Eye-Tracking Feedback for Tuning LLMs. In *ICML 2024 Workshop on LLMs and Cognition*, Vienna, Austria.

Kyuyoung Kim, Ah Jeong Seo, Hao Liu, Jinwoo Shin, and Kimin Lee. 2024. Margin Matching Preference Optimization: Enhanced Model Alignment with Granular Feedback. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13554–13570, Miami, FL.

Reinhold Kliegl, Richard Olson, and Bruce Davidson. 1982. Regression Analyses as a Tool for Studying Reading Processes: Comment on Just and Carpenter's Eye Fixation Theory. *Memory & Cognition*, 10(3):287–296.

Jey Han Lau, Alexander Clark, and Shalom Lappin. 2017. Grammaticality, Acceptability, and Probability: A Probabilistic View of Linguistic Knowledge. *Cognitive Science*, 41(5):1202–1241.

Harrison Lee, Samrat Phatale, Hassan Mansoor, Thomas Mesnard, Johan Ferret, Kellie Lu, Colton Bishop, Ethan Hall, Victor Carbune, Abhinav Rastogi, and Sushant Prakash. 2024. RLAIF vs. RLHF: Scaling Reinforcement Learning from Human Feedback with AI Feedback. In *Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria.

Tianqi Liu, Zhen Qin, Junru Wu, Jiaming Shen, Misha Khalman, Rishabh Joshi, Yao Zhao, Mohammad Saleh, Simon Baumgartner, Jialu Liu, Peter Liu, and Xuanhui Wang. 2025. LiPO: Listwise Preference Optimization through Learning-to-Rank. In *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2404–2420, Mexico City, Mexico.

Yunfei Long, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang. 2017. A Cognition Based Attention Model for Sentiment Analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 462–471, Copenhagen, Denmark.

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. In *Proceedings of the ACM on Human-Computer Interaction*, 3, New York, NY.

Ángela López-Cardona, Carlos Segura, Alexandros Karatzoglou, Sergi Abadal, and Ioannis Arapakis. 2025. Seeing Eye to AI: Human Alignment via Gaze-Based Response Rewards for Large Language Models. In *The 13th International Conference on Learning Representations*, Singapore, Singapore.

Jonathan Malmaud, Roger Levy, and Yevgeni Berzak. 2020. Bridging Information-Seeking Human Gaze and Machine Reading Comprehension. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 142–152, Online.

Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya. 2016. Leveraging Cognitive Features for Sentiment Analysis. In *Proceedings of the 20th Conference on Computational Natural Language Learning*, pages 156–166, Berlin, Germany.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training Language Models to Follow Instructions with Human Feedback. In *Proceedings of Advances in Neural Information Processing Systems*, New Orleans, LA.

Paul Prasse, David Reich, Silvia Makowski, Seoyoung Ahn, Tobias Scheffer, and Lena Jäger. 2023. SP-EyeGAN: Generating Synthetic Eye Movement Data with Generative Adversarial Networks. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, New York, NY. Association for Computing Machinery.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher Manning, and Chelsea Finn. 2023. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, Red Hook, NY.

Keith Rayner. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, 124(3):372–422.

Keith Rayner, Tessa Warren, Barbara Juhasz, and Simon Liversedge. 2004. The Effect of Plausibility on Eye Movements in Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6):1290–1301.

Elizabeth Schotter and Brian Dillon. 2025. A Beginner's Guide to Eye Tracking for Psycholinguistic Studies of Reading. *Behavior Research Methods*, 57(68).

Elizabeth Schotter and Annie Jia. 2016. Semantic and Plausibility Preview Benefit Effects in English: Evidence from Eye Movements. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 42(12):1839–1866.

Gideon Schwarz. 1978. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Ekta Sood, Fabian Kögel, Philipp Müller, Dominike Thomas, Mihai Bâce, and Andreas Bulling. 2023. Multimodal Integration of Human-Like Attention in Visual Question Answering. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 2648–2658, Vancouver, Canada.

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.

Ekta Sood, Simon Tannert, Philipp Mueller, and Andreas Bulling. 2020b. Improving Natural Language Processing Tasks with Human Gaze-Guided Neural Attention. In *Proceedings of Advances in Neural Information Processing Systems*, pages 6327–6341, Online.

Divyank Tiwari, Diptesh Kanojia, Anupama Ray, Apoorva Nunna, and Pushpak Bhattacharyya. 2023. Predict and Use: Harnessing Predicted Gaze to Improve Multimodal Sarcasm Detection. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15933–15948, Singapore, Singapore.

Andrea Tuninetti, Tessa Warren, and Natasha Tokowicz. 2015. Cue Strength in Second-language Processing: An Eye-tracking Study. *The Quarterly Journal of Experimental Psychology*, 68(3):568–584.

Shravan Vasishth, Titus von der Malsburg, and Felix Engelmann. 2013. What Eye Movements can Tell us about Sentence Comprehension. *WIREs Cognitive Science*, 4(2):125–134.

Sheng Wang, Liheng Chen, Jiyue Jiang, Boyang Xue, Lingpeng Kong, and Chuan Wu. 2024. LoRA Meets Dropout under a Unified Framework. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 1995–2008, Bangkok, Thailand.

Alex Warstadt, Amanpreet Singh, and Samuel Bowman. 2019. Neural Network Acceptability Judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641.

Duo Yang and Nora Hollenstein. 2023. PLM-AS: Pre-trained Language Models Augmented with Scanpaths for Sentiment Classification. In *Proceedings of the Northern Lights Deep Learning Workshop*, Tromsø, Norway.

Yang Zhao, Yixin Wang, and Mingzhang Yin. 2024. Ordinal Preference Optimization: Aligning Human Preferences via NDCG. *arXiv*, 2410.04346.

Appendix

A Implementation Details

The pretrained checkpoint was sourced from a publicly available Hugging Face repository. All models were trained on an NVIDIA GeForce RTX 4090 GPU. In all training configurations, we fine-tuned the 7 B instruction-tuned Mistral model with 4 bit weight quantisation; both policy and reference models were quantised. During training, we applied parameter-efficient tuning and therefore updated only the LoRA parameters (rank r=16, α=32, dropout $= 0.1$, no bias). We optimised with the AdamW optimiser under a cosine schedule with a 10-step warm-up, batch size 8, and a maximum sequence length of 512 tokens.

We trained with three random seeds (17, 23, 42). Table 2 lists the hyperparameters explored in the grid search; the final setting was selected based on the lowest validation loss.

Table 2: Hyperparameter grid.

Hyperparameter	Values
Learning rate	2×10^{-6}, 3×10^{-6}, 5×10^{-6}
Weight decay	0.02, 0.03
Training steps	600, 700, 1000, 3000, 4700, 6120
β	0.2, 0.3
α	0.10, 0.05
Number of pairs	20, 30, 40

B Generation of Synthetic Eye Movement Data

To extend the eye movements dataset for training the model in the gaze-augmented DPO setting we generate the synthetic eye movements-while-reading data, particularly we predict the scanpaths for 30 sentence pairs from the CoLA dataset, preprocess the gaze data to extract the event count-based reading measures and train the models on both human and synthetic eye movements data.

Scanpath prediction is the task of mapping a tokenised sentence $x = (w_1, \ldots, w_T)$ to a variable-length sequence of eye-movement events $s = (e_1, \ldots, e_m)$, where each fixation event comprises the index $p_i \in \{1, \ldots, T\}$ of the fixated token. We formalise this as learning a conditional distribution $P(s \mid x; \theta)$, instantiated via autoregressive sequence models or structured prediction frameworks, by minimizing the negative log-likelihood:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{M} \log P(e_i \mid e_{<i}, x; \theta)\,.$$

We used two corpora to train the Eyettention model: CELER (Berzak et al., 2022) and CoLAGaze (introduced above). CELER is a large-scale eye-tracking dataset comprising gaze recordings from 365 participants, including both native (L1) and non-native (L2) English speakers with varying levels of language proficiency and linguistic backgrounds. The participants read a total of 28,548 sentences, randomly sampled from Wall Street Journal (WSJ) newswire text. The dataset provides word-level fixation data, which we used to train the generative model of eye movements.

We generated synthetic fixation sequences for 30 CoLA training sentences using Eyettention, a dual-encoder Transformer for scanpath generation. We evaluated three training configurations: (i) pretrain on CELER (Berzak et al., 2022) and fine-tune on CoLAGaze, (ii) train on CELER only, and (iii) train on CoLAGaze only. Hyperparameters followed the original Eyettention setup. Each configuration used 5-fold cross-validation with the original "new sentence" split.

Training Data	**Fine-Tuning Data**	**Testing Data**	**NLD↓**
CELER	–	$\text{CoLAGaze}_{\text{all}}$	$0.493_{0.074}$
CoLAGaze	–	$\text{CoLAGaze}_{\text{all}}$	$0.487_{0.008}$
CELER	CoLAGaze	$\text{CoLAGaze}_{\text{all}}$	$0.491_{0.008}$
CELER	–	$\text{CoLAGaze}_{\text{ung}}$	$0.491_{0.014}$
CoLAGaze	–	$\text{CoLAGaze}_{\text{ung}}$	$0.484_{0.012}$
CELER	CoLAGaze	$\text{CoLAGaze}_{\text{ung}}$	$0.487_{0.017}$

Table 3: Eyettention training configurations and scanpath quality on CoLAGaze. Normalised Levenshtein Distance (NLD; lower is better) is reported as mean_{SD} over readers. $\text{CoLAGaze}_{\text{all}}$ = all sentences; $\text{CoLAGaze}_{\text{ung}}$ = ungrammatical subset.

Performance was measured on a held-out CoLAGaze subset using normalised Levenshtein distance (NLD) between synthetic and human scanpaths: for each sentence–reader pair we computed the Levenshtein distance, divided it by the maximum scanpath length, and then averaged across readers. We report results for all sentences and for the ungrammatical subset. The three configurations performed similarly; the CoLAGaze-only model was marginally better on both subsets (Table 3). We therefore used this model to generate synthetic scanpaths. From these scanpaths we extracted the same event-count features as for human data, using the identical preprocessing pipeline, and integrated them into the DPO training pipeline.

C Reading Measures

To integrate human cognitive signals into the DPO framework, we extracted a diverse set of eye-tracking measures that capture different aspects of on-line reading behaviour. These measures reflect temporal and spatial dynamics of eye movements and have been shown in psycholinguistic research to be sensitive to lexical and syntactic properties of text. We report them in Table 4.

Reading Measure	Definition
Second pass duration (IQR, mean)	sum of fixation durations when a word is revisited after the first pass reading is complete, before the third pass
Go past time (mean, SD, IQR)	sum of all fixation durations from the first fixation on a word until the reader moves to a word to the right (progresses forward in the text)
First duration (median, IQR)	duration of the first fixation on a word, regardless of whether it was fixated in the first pass or not
Rereading time	duration of all fixation after the first pass
Gaze duration (SD, median, mean)	of all fixation durations on a word during first pass reading (before the eyes leave the word for the first time)
Normalised outgoing regressions count (SD)	number of regressions initiated from a word normalised by the total number of progressive saccades in a sentence
Saccade length (median, SD)	absolute horizontal distance of a saccade, measured in number of characters
Regression rate	proportion of regressions out of total incoming and outgoing saccades
Reading duration	total time spent reading each item, normalized by sentence length
Total fixation duration (SD)	sum of all fixation durations on a word across all passes
First fixation duration (SD, IQR, mean)	duration of the first fixation on a word during the first pass
Saccade duration (SD, IQR)	saccade duration in milliseconds
Normalised saccade duration (IQR)	saccade duration normalized by total reading time
Word in Fixed Context First and Total Fixation Duration (mean)	first and total fixation duration on a word in a fixed context (see Berzak et al., 2018 for more details) normalised by the context overall reading duration
Information Cluster First and Total Fixation Duration (mean, SD)	first and total fixation duration on a word in an information cluster (see Berzak et al., 2017 for more details) normalised by the cluster overall reading duration
Syntactic Cluster Total Fixation Duration (mean)	total fixation duration on a word in a syntactic cluster (see Berzak et al., 2018 for more details) normalised by the cluster overall reading duration

Table 4: Eye-tracking measures employed to augment DPO framework. For each measure we report its definition and the aggregation statistic(s) used to obtain a sentence-level vector (mean/median/SD/IQR).