

Responsible NLP Checklist

Paper title: *EyeMulator: Improving Code Language Models by Mimicking Human Visual Attention*
Authors: *Yifan Zhang, Chen Huang, Yueke Zhang, Jiahao Zhang, Toby Jia-Jun Li, Collin McMillan, Kevin Leach, Yu Huang*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- the authors indicated that the question does not apply to their work
- the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the [Responsible NLP Checklist](#) page at ACL Rolling Review.

A. Questions mandatory for all submissions.

A1. Did you describe the limitations of your work?

This paper has a Limitations section.

A2. Did you discuss any potential risks of your work?

Section 9 (Ethical Considerations) discusses dual-use risks of gaze analysis, particularly the potential for workplace surveillance, and explicitly disclaims per-developer tracking as a use case.

B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?

Section 9 and the Data Availability Statement: we redistribute only aggregated statistics (Beta priors and n-gram counts), not per-participant gaze traces, and rely on the original EyeTrans IRB-approved anonymization protocol.

B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?

Section 2.1 reports the size of the EyeTrans corpus we distill from (120Hz gaze recordings from 27 programmers, 1,565 extracted scan paths). Section 3 reports the CodeXGlue splits used for downstream evaluation (10,300 / 500 for translation, 16,500 for summarization, approximately 1.6k files for completion).

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

Section 3 (Implementation Details) summarizes the training setup; Appendix B.2 (Table 5) provides the full list of shared hyperparameters across all three backbones and all three downstream tasks.

C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?

The Responsible NLP Checklist used at ACL Rolling Review is adopted from NAACL 2022, with the addition of ACL 2023 question on AI writing assistance and further refinements based on ARR practice. ACL 2026 used a subset of ARR checklist form.

Tables 24 report single-run evaluation scores against fixed train/test splits from CodeXGlue under identical hyperparameters and seed initialization across backbones; this is explicitly noted in Section 3. Multi-seed experiments were not conducted due to compute constraints, as acknowledged in Section 8.

D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?

D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

This work reuses the publicly released EyeTrans corpus (Zhang et al., 2024); we did not recruit new human subjects or collect new gaze data. Participant instructions are reported in the original EyeTrans publication.

D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

Same as D1: no new human-subject recruitment or compensation was performed in this work; recruitment and compensation details for the EyeTrans cohort are documented in the original publication.

D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?

Section 9 (Ethical Considerations) and the Data Availability Statement note that EyeTrans data is used under its original IRB-approved consent and anonymization protocols, and we redistribute only aggregated derived artifacts.

D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

Section 9: the original EyeTrans data collection was IRB-approved; we reuse its anonymized data under those terms and release only aggregate statistics.

E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?

E1. If you used AI assistants, did you include information about their use?

Yes. See the "Use of AI Assistants" paragraph in the Ethical Considerations section. AI assistants were used only for minor editorial assistance (phrasing, LaTeX formatting) and not for research ideation, experiment design, implementation, or result analysis; no substantive technical content was AI-generated.