

## Responsible NLP Checklist

Paper title: *GerAV: Towards New Heights in German Authorship Verification using Fine-Tuned LLMs on a New Benchmark*

Authors: *Lotta Kiefer, Christoph Leiter, Sotaro Takeshita, Elena Schmidt, Steffen Eger*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- <sup>N/A</sup> 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?

*Ethical Considerations, page 10*

### 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?

*We neither investigated nor explicitly use any uniquely identifying data. All datasets employed in this work are authorized for research purposes. This applies to both the newly collected Reddit dataset and the Twitter dataset. We ensure reproducibility in compliance with the original licenses, as detailed in Appendix J (page 20) of our paper. Furthermore, we do not collect or store any author metadata. Our research is conducted in accordance with a formal ethical and legal assessment that governs data use and processing, which we strictly adhere to.*

- 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 3, Data Curation*

### C. Did you run computational experiments?

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

*Appendix F: Training setup, hyperparameters and model licences*

- 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 results presented are from a single representative run, which is transparent. Multiple independent runs were not conducted due to computational and time constraints, as usual when fine-tuning LLMs. Therefore, no means, variances, or confidence intervals are reported since they not applicable to our*

---

*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.*

*experiments. To nevertheless assess statistical significance, we employ non-parametric bootstrap tests, which are reported alongside the results.*

**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.?

*(left blank)*

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)?

*(left blank)*

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)?

*(left blank)*

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

*(left blank)*

**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?

*AI assistants were used for spell checking and reformulations of existing content during the paper writing process and for coding assistance during the experimental phase. No new content was generated with AI Assistants.*