

Responsible NLP Checklist

Paper title: *CommonLID: Re-evaluating State-of-the-Art Language Identification Performance on Web Data*

Authors: *Pedro Ortiz Suarez, Laurie Burchell, Catherine Arnett, Rafael Mosquera, Sara Hincapi Monsalve, Thom Vaughan, Damian Stewart, Malte Ostendorff, Idris Abdulmumin, Vukosi Marivate, Shamsuddeen Hassan Muhammad, Atnafu Lambebo Tonja, Hend Al-Khalifa, Nadia GHEZAIEL HAMMOUDA, Verrah Akinyi Otiende, Tack Hwa Wong, Jakhongir Saydaliev, Melika Nobakhtian, Muhammad Ravi Shulthan Habibi, Chalamalasetti Kranti, Carol Muchemi, Khang Nguyen, Faisal Muhammad Adam, Luis Frentzen Salim, Reem Alqifari, Cynthia Jayne Amol, Joseph Marvin Imperial, Ilker Kesen, Ahmad Mustafid, Pavel Stepachev, Leshem Choshen, David Anugraha, Hamada Nayel, Seid Muhie Yimam, Vallerie Alexandra Putra, My Chiffon Nguyen, Azmine Touseh Wasi, Gouthami Vadithya, Rob van der Goot, Lanwenn ar C'horr, Karan Dua, Andrew Yates, Mithil Banger, Yeshil Banger, Hitesh Laxmichand Patel, Shu Okabe, Fenal Ashokbhai Ilasariya, Dmitry Gaynullin, Genta Indra Winata, Yiyuan Li, Juan Pablo Martnez, Amit Agarwal, Ikhlusal Akmal Hanif, Raia Abu Ahmad, Esther Adenuga, Filbert Aurelian Tjiaranata, Weerayut Buaphet, Michael Anugraha, Sowmya Vajjala, Benjamin L Rice, Azril Hafizi Amirudin, Jesujoba Oluwadara Alabi, Srikant Panda, Yassine Toughrai, Bruhan Kyomuhendo, Daniel Ruffinelli, Akshata, Manuel Goulo, Ej Zhou, Ingrid Gabriela Franco Ramirez, Cristina Aggazzotti, Konstantin Dobler, Jun Kevin, Quentin PAGES, Nicholas Andrews, Nuhu Ibrahim, Mattes Ruckdeschel, Amr Keleg, Mike Zhang, Casper Rufaro Muziri, Saron Samuel, Sotaro Takeshita, Kun Kerdthaisong, Luca Foppiano, Rasul Dent, Tommaso Green, Ahmad Mustapha Wali, Kamohelo Makaaka, Vicky Feliren, Inshirah Idris, Hande Celikkanat, Abdulhamid Abubakar, Jean Maillard, Benot Sagot, Thibault Clrice, Kenton Murray, Sarah K. K. Luger*

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?

Limitations

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?

Limitations

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.

- 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?

Appendix B

C. Did you run computational experiments?

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

Section 5

- 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?

Section 4, Appendix D

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.?

Appendix E

- 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)?

Section 3

- 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 3, Limitations

- 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?

(left blank)