

Responsible NLP Checklist

Paper title: *Do ImageText Metrics Respect Semantic Invariances?*

Authors: *Amit Agarwal, Hitesh Laxmichand Patel, Meizhu Liu, Jyotika Singh, Karan Dua, Hansa Meghwani, Matthew Rowe, M. Avendi, Yassi Abbasi, Tao Sheng, Sujith Ravi, Dan Roth*

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?

Discussed in Ethical Considerations (misuse risk + mitigation; also notes potential bias/offensive content in public datasets).

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?

Ethical Considerations. (We use established public datasets and do not collect new personal data; we mitigate exposure by not redistributing images, only identifiers/derived scores).

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 3.1 (Datasets and Curation) and Appendix A.2 (Datasets, Curation, and Balance) describe sources/splits, inclusion criteria, bins/taxonomy, and balance/disclosure. Human-study sample sizes are in Appendix A.11.

C. Did you run computational experiments?

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

Setup/parameters for perturbations and evaluation protocol are in Section 3.23.5 (perturbation parameters, bootstrap/CIs, calibration sweep + constraint), with evaluator/judge configuration details in Appendix A.8.

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?

Descriptive statistics are reported throughout Results (Section 4) using medians and 95% BCa bootstrap CIs, with reporting conventions in the methodology/results text and figure captions.

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.

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.?

The annotation questions/options (acceptability + preference) are specified in Appendix A.11 (Human validation of perturbations).

- 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 model was used solely for drafting and editorial assistance, while all core ideas, analyses, and contributions were developed by the authors