Responsible NLP Checklist

Paper title: *MMAG: Multimodal Learning for Mucus Anomaly Grading in Nasal Endoscopy via Semantic Attribute Prompting*
Authors: *Xinpan Yuan, Mingzhu Huang, Liujie Hua, JianuoJu, XuZhang*

How to read the checklist symbols:

☑ the authors responded 'yes'

☒ the authors responded 'no'

N/A 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?
The framework may miss or misclassify mucus in extremely low-light areas, underestimating rhinitis severity.

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

☑ B1. Did you cite the creators of artifacts you used?
The creators of the polyp-related datasets used (e.g., CVC_ClinicDB, CVC_ColonDB, Hyper-Kvasir) have been cited via references in the "4.1 Experimental Setups" section.

☑ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
Public polyp datasets (CVC_ClinicDB, etc.) have their original licenses cited in "4.1 Experimental Setups" per creators terms; the self-curated nasal dataset is authorized by ethics approval (LLY PJ2025091-01) with participant consent.

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
The paper confirms existing artifacts are used as intended, specifies self-created ones for rhinitis research (compatible with access rules), and notes curated data is anonymized.

☑ 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?
The paper states the curated nasal dataset was anonymized (per "Ethical Considerations") and checked for identifiers/offensive content via expert review (none found).

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?

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.

The paper documents artifacts (CND/PUS dataset, polyp datasets, CLIP model) in "4.1 Experimental Setups", covering domains (nasal endoscopy/polyp medical imaging), data sources, and demographic scope (hospital patients).

☑ 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?
The paper reports relevant statistics (e.g., number of examples, train/test/dev splits) for used/created data (CND/PUS, polyp datasets) in "4.1 Experimental Setups".

☑ **C. Did you run computational experiments?**

N/A C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
Reports infrastructure (NVIDIA Tesla V100 GPUs) and training details, but not model parameters or total GPU hours.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
The paper discusses experimental setup in "Implementation Details", including hyperparameter search (AdamW optimizer, learning rate 0.0001, 100 - epoch training, batch size 16) and best - found hyperparameters.

☒ 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?
Despite dataset splitting details (7:2:1 ratio), only final metrics are shown, with no descriptive stats (e.g., error bars) or clarity on result type (single run/mean/max).

☒ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?
The paper mentions CLIP-ViT/L14 and AdamW optimizer but lacks existing package details (implementation, versions, parameters).

☑ **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 currently referenced document does not report the full text of participant instructions (e.g., screenshots) or risk disclaimers, nor does it mention such information in supplements, even though its main contribution involves human subjects.

☒ 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)?
It does not report participant recruitment methods (e.g., specific crowdsourcing platforms, student details) or payment information, nor does it discuss payment adequacy based on participants demographics (e.g., country of residence).

☒ 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)?
It only mentions details such as the datasets training quantity and source hospital. It does not discuss whether and how consent was obtained from individuals whose data is being used/curatedfor example, it does not explain if relevant parties were informed about how the data would be used.

☑ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Approved by institutional ethics committee (Approval No. LLYPJ2025091 - 01), with written informed consent from participants/guardians.

☒ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
Only basic dataset info is introduced, without reporting annotator demographic/geographic characteristics, stating protected info inclusion, or providing data statements and explanations on human subject features.

☑ **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 (e.g., ChatGPT) were used only for language polishing, not affecting scientific content or results.