

## Responsible NLP Checklist

Paper title: *Towards the First NLP Benchmark for Ladin - an Extremely Low-Resource Language*

Authors: *Ulin Nuha, Adam Jatowt*

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 work does not involve any direct risks to participants, as the data used is primarily publicly available datasets. No personal or sensitive information was involved in the data collection process, and the task itself posed no harm or risk to participants.*

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

#### B1. Did you cite the creators of artifacts you used?

*Section 3: This section contains citations to the creators of datasets used in our research*

#### B2. Did you discuss the license or terms for use and/or distribution of any artifacts?

*The datasets and resources used in this work are sourced under open licenses or with explicit permission from the original owners.*

#### 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)?

*We specify that all datasets is under an open access and get the permit from the author*

#### 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 did not collect or use data that contains personally identifying information or offensive content. The datasets used in this research consist of paired Italian-Ladin sentence pairs and Italian monolingual data, and no personally identifiable information was included.*

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

*We provide documentation of the datasets used in our research, including the coverage of domains, languages (Ladin and Italian), and linguistic phenomena*

---

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.

- 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 and 4*

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

- C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?  
*Section 3*

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?  
*: Section 3*

- 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*

- 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?  
*We used existing packages with their default settings for evaluation (e.g., ROUGE, BLEU, and others). As a result, no specific implementation, model, or parameter settings were modified from their default configurations.*

**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 task consisted of simple manual and evaluation translation from Italian to Ladin.*

- 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)?  
*The translations were primarily provided by a native Ladin speaker and not through a formal annotation process.*

- 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)?  
*We already got the consent from the annotator.*

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?  
*The data used in our research was not collected through formal recruitment or a structured protocol requiring ethics review board approval. The translations were done manually by a friend from the region, and we also used publicly available datasets. Consent was obtained from the authors of the datasets for their use in our research, and no formal ethics review board approval was required in this context.*

- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?  
*We did not report specific demographic or geographic characteristics of the annotator population since the focus is only on Ladin (Val Badia variant) native speaker*

**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?

*We did not explicitly mention the use of AI assistants, such as ChatGPT, in the paper, but they were used to assist in troubleshooting and resolving errors during the research process.*