

## Responsible NLP Checklist

Paper title: *ServImage: An Image Generation and Editing Benchmark from Real-world Commercial Imaging Services*

Authors: *Fengxian Ji, Jingpu Yang, Zirui Song, Lang Gao, Junhong Liang, Zhenhao Chen, Jinghui Zhang, Xiuying Chen*

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?

*Sec. Ethics Statement*

### 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 state that any released data should be anonymized and stripped of PII (with sensitive content restricted/redacted), but we do not detail a concrete screening procedure for PII or offensive content.*

- 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?

*Sec. 3.2*

### C. Did you run computational experiments?

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

*We do not describe a hyperparameter search procedure; we only report the fixed hyperparameter settings used for training.*

- 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?

*We report single point estimates aggregated over the full benchmark (e.g., Table 2 and Fig. 5) and do not include error bars, confidence intervals, or multi-run statistics.*

*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.?

*While the paper describes the annotation workflow and evaluation setup at a high level, it does not include the full text of participant/annotator instructions, interface screenshots, or explicit risk disclaimers.*

- 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 paper mentions that the data comes from real commercial orders on two crowdsourcing platforms (Epmk.com, Zbj.com), but it does not report how participants/annotators were recruited and compensated, nor does it discuss whether the compensation is sufficient in the context of the country/population they are from.*

- 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)?

*The paper does not explicitly discuss whether or how consent was obtained from individuals whose data were used or curated, nor whether participants were informed about how their data would be used.*

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

*The paper does not report approval from, or exemption by, an ethics review board for the data collection protocol.*

**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, including ChatGPT and Claude, were used during the preparation of this manuscript. Specifically, they were employed to assist with language polishing, grammar correction, and generating template code for data visualization scripts. All AI-generated outputs were critically reviewed, edited, and verified by the authors to ensure accuracy and original contribution.*