

Paper Validation Report for OSA



Correspondence: Yes

Percentages: 90.0%

Conclusion:

The codebase implements the methods, models, and algorithms described in the paper's experiments section: OSA automates README and docstring generation, repository analysis, and uses LLMs and defined prompt/templates as described. Classes for loading prompts, running evaluations, structuring reports, and comparing results (e.g., via BERT scores, Flesch-Kincaid, conciseness) are present and align with what is outlined in the experiments. The description and code reveal validation sets, evaluation metrics, and procedures for assessing README and docstring quality. However, exact dataset splits, some metric calculation scripts, and full reproduction of tables/figures may not be end-to-end executable directly from the code modules alone, requiring external benchmarking or minor adaptation. Overall, the core experimental claims are supported, but slight gaps in explicit script-orchestration and benchmarks prevent a perfect score.