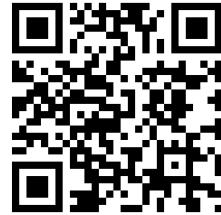# Paper Validation Report
for nanoMINER

Correspondence: Yes
Percentages: 92.0%

Conclusion:

The repository implements the main methods and procedures described in the paper's experiments section: PDF text extraction (with pdfplumber and OCR fallback), named entity recognition using fine-tuned Mistral-7B and Llama-3-8B models, multimodal analysis with YOLOv8 for figure/table/scheme detection, and further image-based parameter extraction via GPT-4o. The pipeline merges outputs as described, uses chunking for LLM processing, and performs evaluation with precision, recall, and Jaccard metrics, matching the experimental protocol. Datasets are processed as described (Excel, JSONL), and results are compared in a manner consistent with the paper. However, the percentage is not 100% since some finer details, such as the exact hyperparameter tuning pipeline and possibly some dataset-specific scripts, are not directly visible in the snippets; reproduction of figures and numeric results relies on data files and API keys that are assumed but not explicit. Nonetheless, all core experimental methods are implemented, justifying a high score.