



RepoAgent: An LLM-Powered Open-Source Framework for Repository-level Code Documentation Generation

Qinyu Luo*, Yining Ye*, Shihao Liang, Zhong Zhang, Yujia Qin, Yaxi Lu, Yesai Wu, Xin Cong, Yankai Lin, Yingli Zhang, Xiaoyin Che, Zhiyuan Liu, Maosong Sun



Introduction

Developers typically spend approximately 58% of their time on program comprehension, and high-quality code documentation plays a significant role in reducing this time. However, maintaining code documentation also consumes a considerable amount of time, money, and human labor, and not all projects have the resources or enthusiasm to prioritize documentation as their top concern.

To address this, we introduced **RepoAgent**, an automated framework for seamless documentation **generation, updating, and maintenance**.

	Total	Human	Model	Win Rate
Transformers	150	45	105	0.70
LlamaIndex	150	13	137	0.91

Table 1: Result of human preference test on human-authored and model-generated code documentation.

Evaluation Result

Repository	Llama-2-7b	Llama-2-70b	gpt-3.5-turbo	gpt-4-0125
unoconv	0.0000	0.5000	1.0000	1.0000
simdjson	0.4298	0.6336	1.0000	0.9644
greenlet	0.5000	0.7482	0.9252	0.9615
code2flow	0.5145	0.6171	0.9735	0.9803
AutoGen	0.3049	0.5157	0.8633	0.9545
AutoGPT	0.4243	0.5611	0.8918	0.9527
ChatDev	0.5387	0.6980	0.9164	0.9695
MemGPT	0.4582	0.5729	0.9285	0.9911
MetaGPT	0.3920	0.5819	0.9066	0.9708

Table 2: Accuracy of identifying parameters with different LLMs.

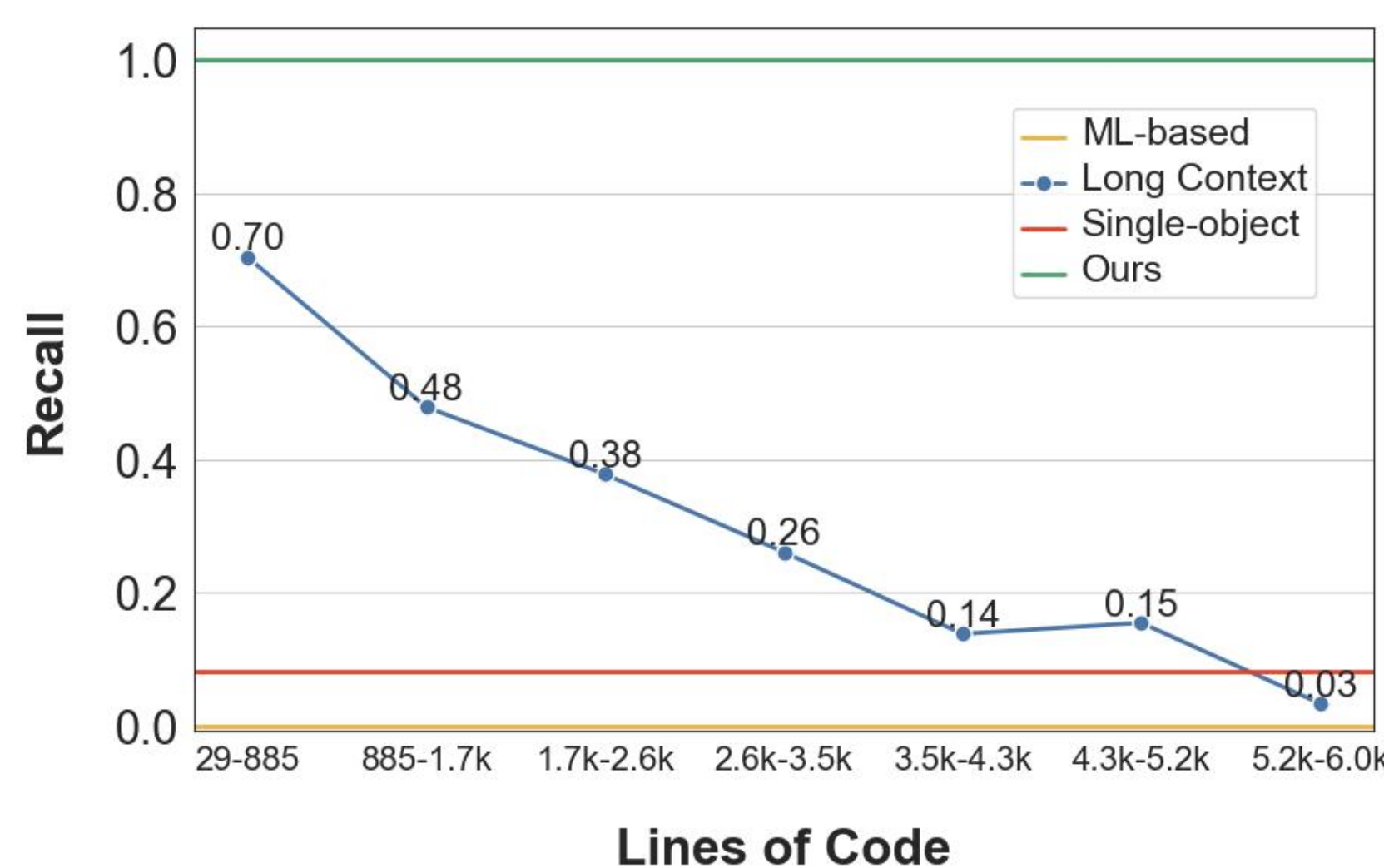


Fig 1: Callee & Callers identification rate.

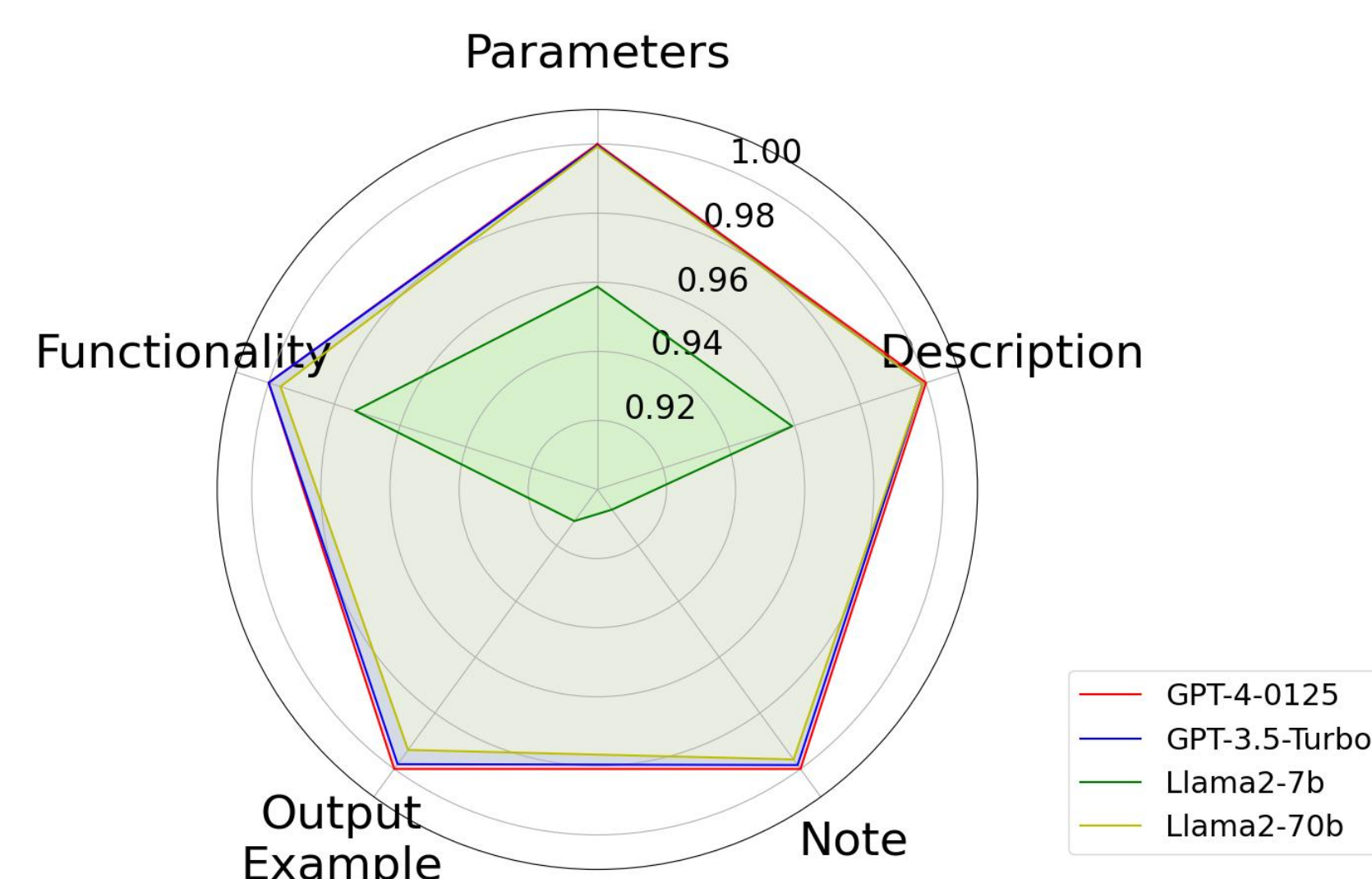
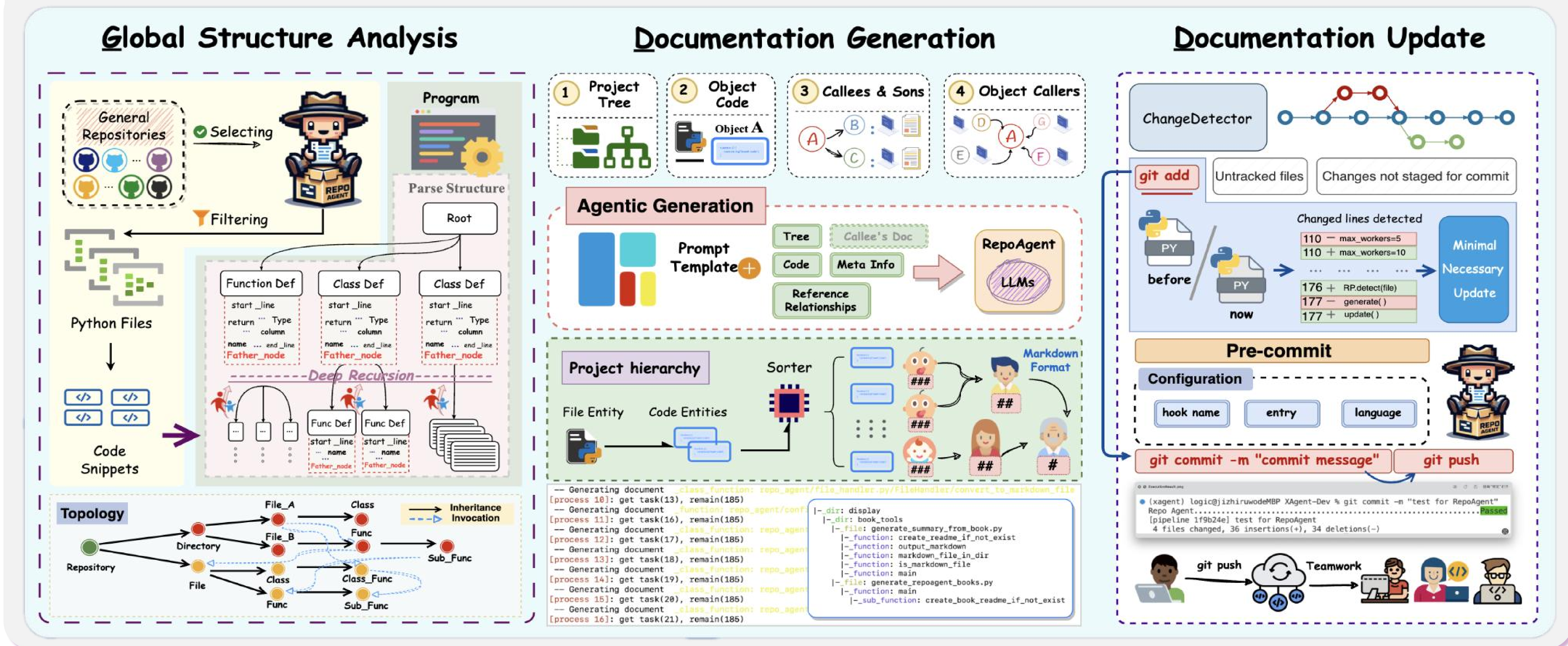


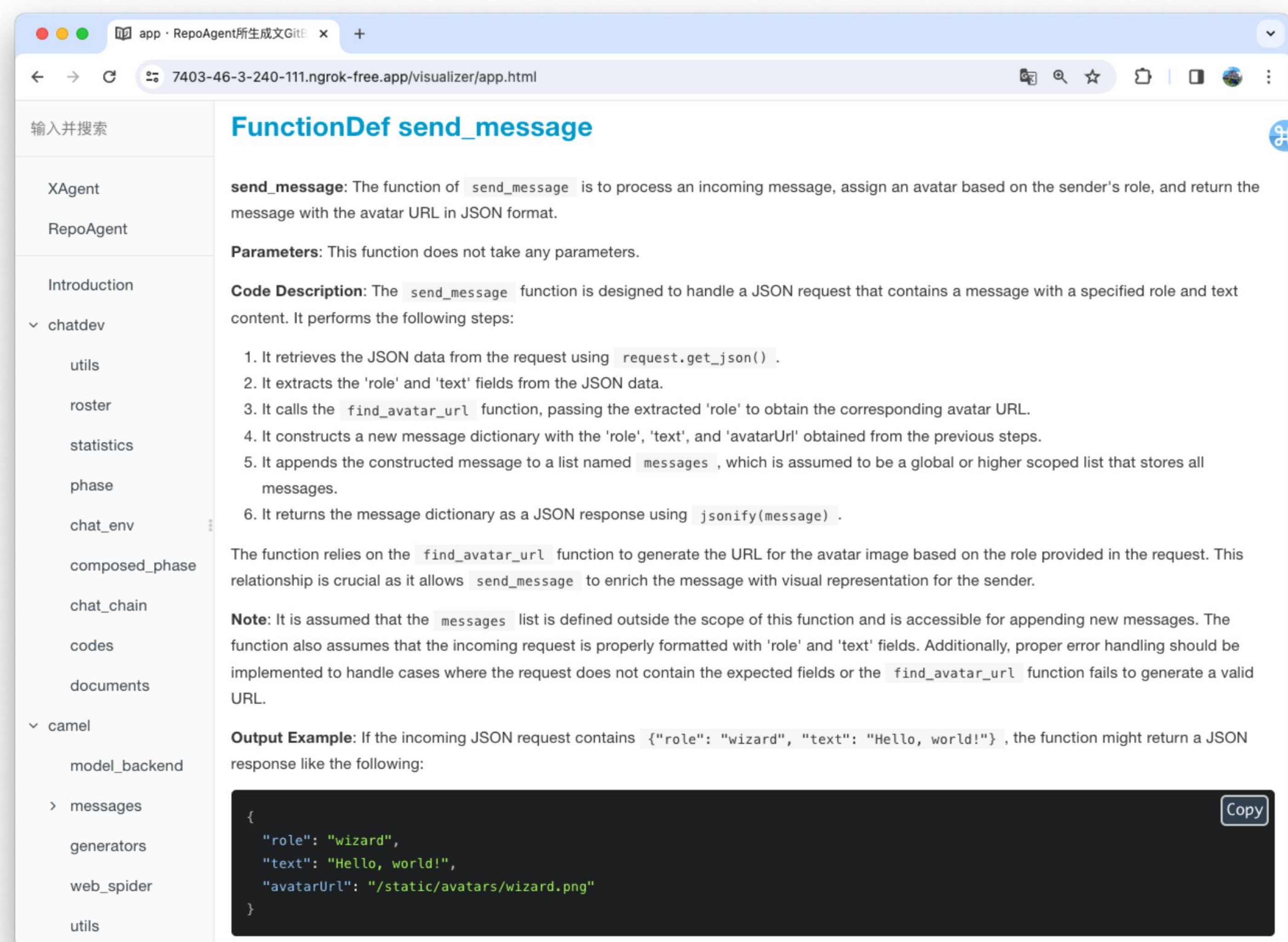
Fig 2: Format align accuracy of LLMs.

Methodology



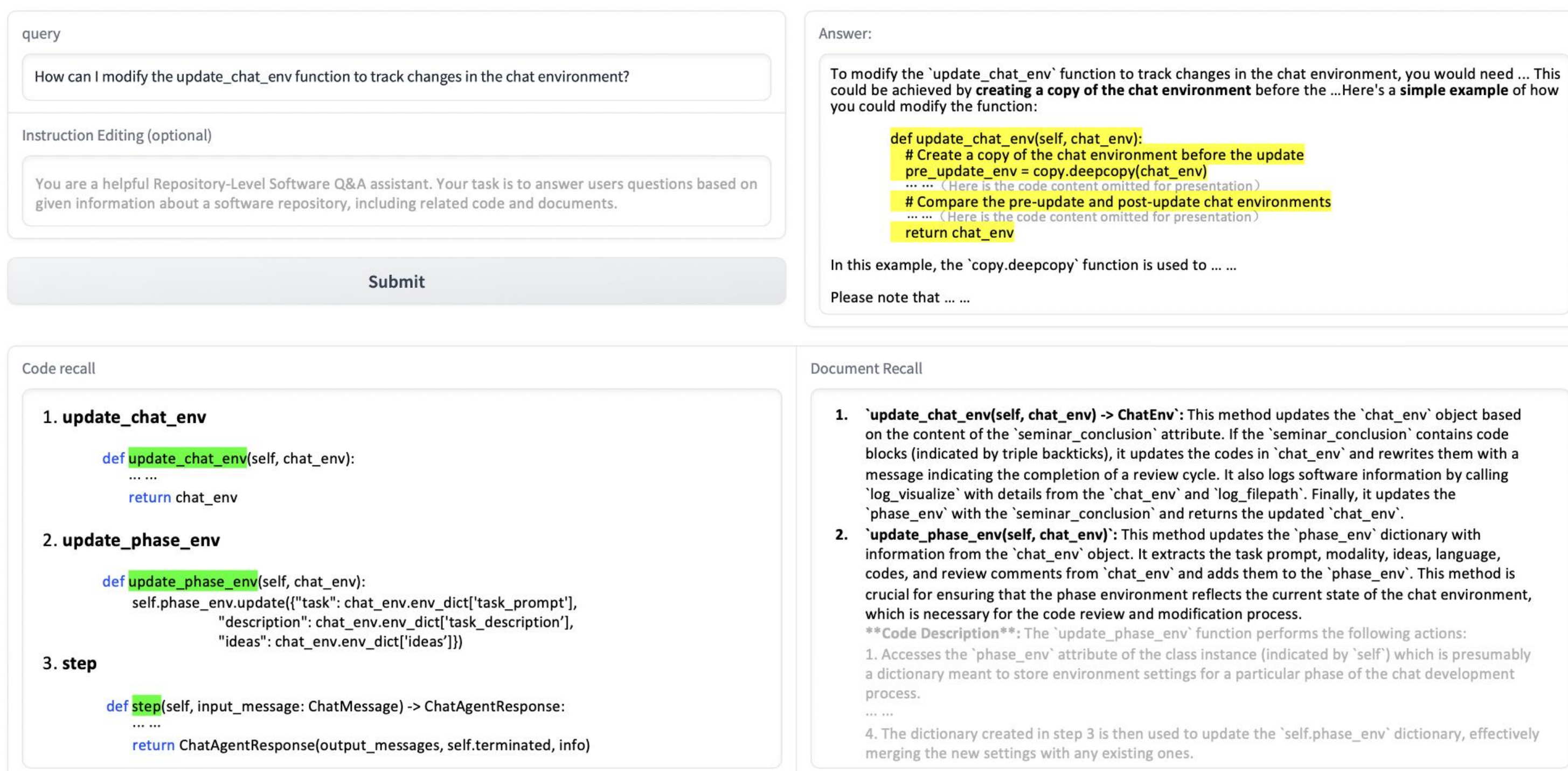
RepoAgent consists of three key stages: **Global structure analysis**, **Documentation generation**, and **Documentation update**.

Final Output



One more thing: Chat With Repo

Automatic Q&A for Issues and Source Codes



Chat With Repo seamlessly integrates advanced retrieval-augmented generation (RAG) techniques to enable automated Q&A, code explanations, and various other downstream applications. By using documentation as a bridge, it maps the search space from natural language to code, enhancing search accuracy and enabling precise repository-level interactions.