

УДК 004.41

РАЗРАБОТКА АГЕНТНОГО OPEN SOURCE РЕШЕНИЯ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА РЕПОЗИТОРИЕВ НА GITHUB С ИСПОЛЬЗОВАНИЕМ LLM

Ульянова Е.А. (ИТМО), Гетманов А.Н. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Никитин Н.О. (ИТМО)

Введение. По статистике за 2024 год количество проектов с открытым исходным кодом, представленных на GitHub, выросло по сравнению с прошлым годом на 98%. Более того, в них увеличилась частота использования генеративного искусственного интеллекта [1]. В повышении удобства их использования и модификации важную роль играет процесс документирования (в том числе добавления docstrings) и написания readme-файла, содержащего исчерпывающую информацию о целях проекта, его основных функциях и способах его использования. На данный момент уже существуют готовые open source решения для автоматизации данных процессов. Так, в качестве примера могут быть рассмотрены readmeai (имеет более 1,5 тысяч звёзд на GitHub), используемый для генерации readme-файлов, и doc-comments-ai – для документации. Оба проекта продолжают развиваться (последние изменения были внесены не раньше января 2025 года), однако по-прежнему имеют ряд недостатков.

Основная часть. OSA (Open-Source-Advisor) – находящееся в процессе разработки решение, предоставляющее возможность получения pull request со всеми предлагаемыми изменениями: обновлённым программным кодом с комментариями (на английском или русском языках), README и документацией (на английском), сформированными с использованием LLM (доступно несколько моделей на выбор) на основе сбора и обработки данных о репозитории и содержимом его файлов, а также тексте научной статьи при её наличии.

Выводы. К февралю 2025 года было проведено тестирование промежуточной версии OSA на репозиториях лабораторий ИТМО, а также одной зарубежной команды, проанализированы его результаты, на основе которых выделены возможности расширения функциональности проекта.

Список использованных источников:

1. GitHub Staff Octoverse: AI leads Python to top language as the number of global developers surges / GitHub Staff [Электронный ресурс] // github.blog : [сайт]. — URL: <https://github.blog/news-insights/octoverse/octoverse-2024/> (дата обращения: 10.02.2025).
2. Dvivedi S. S. et al. A comparative analysis of large language models for code documentation generation // Proceedings of the 1st ACM International Conference on AI-Powered Software. – 2024. – С. 65-73.