

## РАЗРАБОТКА ПРОГРАММНЫХ МОДУЛЕЙ, РАСШИРЯЮЩИХ ВОЗМОЖНОСТИ ОТКРЫТОГО ФРЕЙМВОРКА OSA

Ульянова Е. А.<sup>1</sup>

Научный руководитель – м.н.с. Гетманов А.Н.<sup>1</sup>

<sup>1</sup>Университет ИТМО

Oulyanova.ekaterina.4@gmail.com

Работа выполнена в рамках темы НИР № 625134 «Исследование и разработка фронтальных методов искусственного интеллекта и их приложений».

### Введение

Платформы для совместной разработки, такие как GitHub, GitLab и Gitverse, стали фундаментом современного программирования. Масштаб активности иллюстрирует статистика GitHub: только за 2025 год здесь создано более 121 млн новых репозиторий, а число участников в открытых проектах превысило 1,12 млрд [1]. Однако стремительный рост числа проектов обостряет проблему их качества. Значительная часть репозиторий страдает от неполной документации, неочевидной структуры и несоблюдения лучших практик, что напрямую влияет на воспроизводимость исследований и возможность повторного использования кода. Фреймворк с открытым кодом OSA (Open Source Advisor) решает эти проблемы, предлагая мультиагентный подход к автоматизированному улучшению репозиторий. Несмотря на наличие базовой функциональности (генерация README, CI/CD-файлов, лицензий), существовала потребность в расширении возможностей инструмента для работы со структурой проекта и специфическими форматами, а также в оптимизации использования языковых моделей (LM). Целью данной работы стала разработка и интеграция в OSA трёх ключевых модулей: модуля интеллектуальной реорганизации структуры репозитория (с поддержкой популярных языков, используемых в 80% новых репозиторий [1], а также специализированных форматов, таких как LaTeX), модуля анализа Jupyter-ноутбуков (.ipynb) и модуля задача-ориентированного выбора языковых моделей.

### Основная часть

В ходе работы были спроектированы, реализованы, протестированы и интегрированы три модуля, расширяющие функциональность OSA. Модуль реорганизации структуры включает синтаксическую проверку, устранение ошибок через LM, перемещение файлов с каскадным обновлением импортов и повторную верификацию с возможностью отката. Также была добавлена экспериментальная функция генерации названия репозитория. Модуль анализа Jupyter-ноутбуков выполняет пакетную проверку всех .ipynb-файлов, выявляя синтаксические ошибки и отклонения от стандартов оформления. По результатам генерируется структурированный отчёт, а однозначные ошибки исправляются автоматически. Модуль интеллектуального выбора языковых моделей заменяет единую LLM на механизм дифференцированного выбора модели в зависимости от задачи (docstrings, README, отчёты). Для обоснования выбора моделей по умолчанию проведены эксперименты по генерации docstrings и README с использованием различных SLM и промптов. Оценка выполнялась по метрикам, апробированным в аналогичных исследованиях: BERTScore, читаемость (Flesch–Kincaid), лаконичность, completeness, helpfulness, truthfulness и pass@1 [2-3].

### Выводы

Разработанные модули существенно расширяют функциональность OSA. Реорганизация структуры и анализ `.ipynb` автоматизируют важные задачи подготовки репозитория к публикации. Интеллектуальный выбор моделей и адаптация промптов оптимизируют использование вычислительных ресурсов без существенной потери качества. При этом для ряда метрик наблюдалось его улучшение: лаконичность сгенерированной документации возросла на 14%, а читаемость (Flesch–Kincaid) — на 3% по сравнению с LLM.

### Литература

1. GitHub Octoverse 2025 [Электронный ресурс]. – Режим доступа: <https://itsfoss.com/news/github-octoverse-2025/> (Дата обращения 14.02.2026).
2. Nikitin N. et al. A Multi-Agent LLM Tool for End-to-End Scientific Repository Improvement [Электронный ресурс]. – Режим доступа: <https://openreview.net/forum?id=P3nFmjydO> (дата обращения: 14.02.2026).
3. Yang D. et al. Docagent: A multi-agent system for automated code documentation generation //Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations). 2025. С. 460-471.