**Разработка инструмента с открытым исходным кодом для автоматизированного машинного обучения на основе больших языковых моделей**

**Лапин Алексей Александрович** (ИТМО)

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

**Введение.** Современный мир сталкивается со стремительным ростом объемов данных, что делает ручной анализ неэффективным и зачастую невозможным. Машинное обучение (МО) становится ключевым инструментом для извлечения ценной информации и принятия обоснованных решений в различных областях. Однако эффективное использование моделей МО требует наличия квалифицированных специалистов – аналитиков данных, спрос на которых значительно превышает предложение. Этот дефицит кадров является существенным препятствием для широкого внедрения технологий анализа данных. Процесс создания моделей МО включает трудоемкие и повторяющиеся операции, такие как предобработка данных, выбор и обучение моделей, настройка гиперпараметров и оценка качества, что приводит к значительным временным издержкам и повышает риск ошибок.

Параллельно развитие больших языковых моделей (БЯМ), ознаменовало качественный скачок в сфере обработки естественного языка, демонстрируя впечатляющие способности в понимании, генерации текста, написании кода и логических рассуждениях. Интеграция БЯМ с фреймворками автоматизированного машинного обучения (АвтоМО) представляет собой перспективное направление для снижения барьеров в использовании МО. БЯМ могут обеспечить интуитивный интерфейс на естественном языке, автоматизировать формулировку задачи и подготовку данных, а также помочь в объяснении сложных концепций. Таким образом, гибридные БЯМ-АвтоМО системы способны расширить доступ к технологиям МО для широкой аудитории, не обладающей глубокой экспертизой в области анализа данных.

Целью данной работы является сокращение времени создания и обучения моделей машинного обучения специалистами, имеющими ограниченные навыки разработки приложений в области машинного обучения.

**Основная часть.** Для достижения поставленной цели в рамках работы был разработан и реализован инструмент FEDOT.LLM, использующий синергию больших языковых моделей и классического фреймворка АвтоМО FEDOT [1]. Архитектура FEDOT.LLM основана на концепции гибридного интеллекта, где БЯМ используются для когнитивных функций, таких как анализ постановки задачи, генерация кода и формирование отчетов, а фреймворк FEDOT – для оптимизации и поиска сложных структур конвейеров.

Ключевым элементом системы является многоагентная архитектура, представленная на рисунке. FEDOT.LLM включает в себя три специализированных агента:

* Агент Оркестратор анализирует запрос пользователя и маршрутизирует его соответствующему специализированному исполнителю.
* Агент Исследователь отвечает за поиск и предоставление экспертной информации, связанной с фреймворком FEDOT и методологиями машинного обучения. Для решения проблемы «галлюцинаций» и работы с актуальной, узкоспециализированной документацией агент использует улучшенный подход Retrieval-Augmented Generation (RAG) [2] на основе алгоритма Self-RAG [3]. Этот алгоритм включает этапы поиска релевантных документов, оценки их релевантности, генерации ответа на основе найденного, проверки ответа на обоснованность и полезность, и, при необходимости, переформулировки вопроса или повторной генерации.
* Агент АвтоМО специализируется на задачах построения и обучения моделей машинного обучения с использованием фреймворка FEDOT. Он отвечает за анализ задачи, генерацию конфигураций FEDOT, создание исполняемого Python-кода, его тестирование и автоматическое исправление выявленных проблем, а также формирование отчетов с интерпретацией результатов. Для преодоления «галлюцинаций» БЯМ при генерации кода для фреймворка FEDOT агент использует библиотечные функции, инкапсулирующие реальный программный интерфейс FEDOT и предоставляющие расширенные сообщения об ошибках.

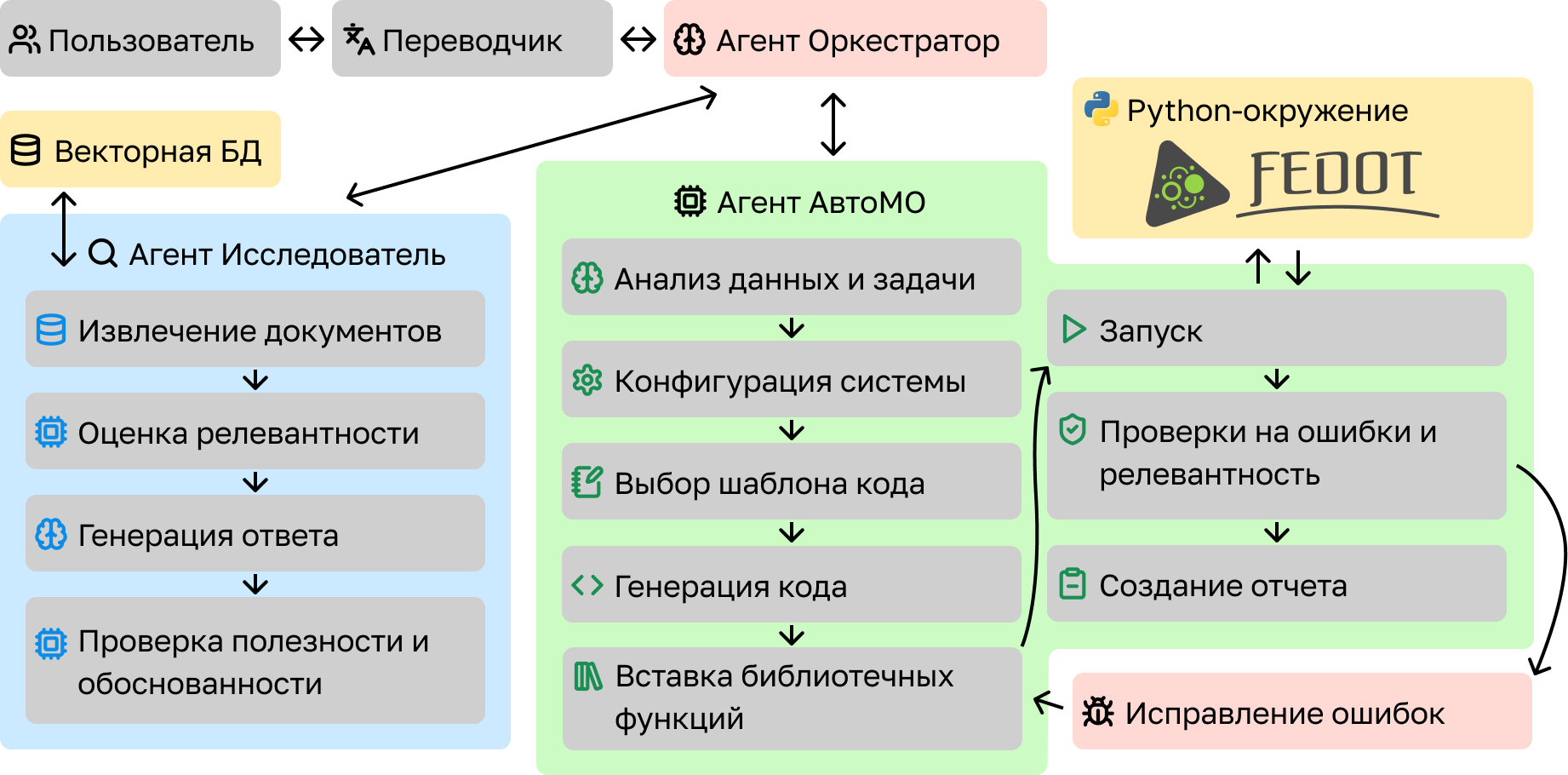


Рисунок. Многоагентная архитектура разработанного инструмента

Система FEDOT.LLM ориентирована на пользователей разного уровня подготовки, предоставляя интуитивный веб-интерфейс для начинающих пользователей и программный интерфейс (API) для экспертов. Взаимодействие осуществляется через естественный язык, что значительно снижает порог входа.

Эффективность разработанного инструмента была подтверждена в ходе трех экспериментов.

1. Сравнительный анализ на задачах Kaggle: Инструмент был протестирован на восьми задачах классификации и регрессии с платформы Kaggle. Результаты показали, что FEDOT.LLM демонстрирует сопоставимую, а в 4 из 8 задачах превосходящую эффективность по сравнению с конкурирующими БЯМ-АвтоМО решениями, такими как AutoKaggle [4] и AIDE [5]. Кроме того, предложенный инструмент стабильно показывает эффективность выше 25-го процентиля решений людей и в 5 из 8 задачах превышает 50-й процентиль.
2. Апробация на прикладных задачах: Инструмент был успешно интегрирован в проект коллег, разрабатывающих ассистента для химика ChemCoScientist, где решал задачи предсказания формы наноматериалов. Это подтвердило способность инструмента справляться с предметно-ориентированными проблемами и возможность его интеграции во внешние системы.
3. Оценка времени разработки: Эксперимент по оценке сокращения времени разработки систем МО показал кардинальное сокращение как временных затрат, так и требуемой квалификации. Если традиционная ручная разработка требовала около 2,5 часов от квалифицированного разработчика, а использование классического АвтоМО-фреймворка – около 0,5 часа от разработчика с базовыми навыками, то с FEDOT.LLM пользователю без специальных знаний потребовалось менее 1 минуты на ввод описания задачи, при этом все последующие шаги были выполнены автономно.

**Выводы.** В ходе выполнения работы была успешно достигнута цель – сокращение времени создания и обучения моделей машинного обучения специалистами с ограниченными навыками разработки. Разработанный инструмент FEDOT.LLM эффективно преодолевает барьеры, связанные с высокой трудоемкостью и требованиями к компетенциям в области МО, способствуя его демократизации.

Ключевые результаты включают всесторонний анализ предметной области, разработку многоагентной архитектуры, объединяющей БЯМ и классический АвтоМО, а также создание функционального программного средства с интуитивным пользовательским интерфейсом. Экспериментальная оценка подтвердила высокую эффективность FEDOT.LLM, его конкурентоспособность по сравнению с существующими решениями и значительное сокращение временных затрат и требований к квалификации пользователей.

Исходный код разработанной системы FEDOT.LLM доступен в официальных репозиториях AIM.club под лицензией 3-Clause BSD. Результаты работы были представлены на XIV Конгрессе молодых ученых ИТМО в секции «Технологии искусственного интеллекта», где была выиграна номинация «За лучший доклад молодого ученого».

**Список использованных источников**:

1. Nikitin N. O. et al. Automated evolutionary approach for the design of composite machine learning pipelines // Future Generation Computer Systems. – 2022. – Т. 127. – С. 109-125.
2. Google Cloud: Retrieval-Augmented Generation [Электронный ресурс]. – Режим доступа: https://cloud.google.com/use-cases/retrieval-augmented-generation?hl=en, свободный (дата обращения: 18.04.2025).
3. Asai A. et al. Self-rag: Learning to retrieve, generate, and critique through self-reflection //The Twelfth International Conference on Learning Representations. – 2023.
4. Li Z. et al. Autokaggle: A multi-agent framework for autonomous data science competitions // arXiv preprint arXiv:2410.20424. – 2024.
5. Jiang Z. et al. Aide: Ai-driven exploration in the space of code //arXiv preprint arXiv:2502.13138. – 2025.
6. Gu Y. et al. Large language models for constructing and optimizing machine learning workflows: A survey //arXiv preprint arXiv:2411.10478. – 2024.