

АВТОМАТИЧЕСКАЯ ПЕРВИЧНАЯ ПРОВЕРКА СТРУКТУРЫ И СОДЕРЖАНИЯ ЛАБОРАТОРНЫХ РАБОТ НА ОСНОВЕ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Лабуткин И.А.¹

Научный руководитель – кандидат технических наук Кугаевских А.В.¹

¹Университет ИТМО

i.labutkin@outlook.com

Введение

В современном образовании наблюдается стремительная интеграция ИИ: согласно данным международных опросов [1], большинство представителей академического сообщества используют ИИ-инструменты, а рынок систем автоматизированной проверки демонстрирует ежегодный устойчивый рост. Однако, несмотря на активное развитие технологий оценивания, существующие подходы к автоматизации проверки письменных работ имеют ряд ограничений. В результате анализа исследований в рассматриваемой предметной области было выявлено, что применение методов машинного и глубокого обучения требует наличие размеченных наборов данных, что делает их непригодными при проверке работ по новым или узкоспециализированным дисциплинам, а более современные подходы на основе больших языковых моделей показывают высокие результаты [2], но критически зависят от наличия примеров проверки работ и заранее формализованных критериев.

Ключевой проблемой является отсутствие универсальных решений, способных автоматически выделять требования к проверке непосредственно из текста задания по широкому кругу дисциплин, не опираясь при этом на готовые примеры. Без решения этой задачи масштабирование систем автоматической проверки на различные предметы является трудоёмким.

Цель работы: повышение качества автоматизированной проверки лабораторных работ по профильным дисциплинам направления "Программная инженерия" для сокращения временных затрат преподавателя на больших потоках (больше 50) студентов за счёт внедрения мультиагентной системы на базе больших языковых моделей

Основная часть

Для решения выделенных проблем была разработана архитектура системы, где процесс проверки декомпозирован на этапы, выполняемые специализированными LLM-агентами. В отличие от использования единой модели, такая система позволяет гибко обрабатывать контекст и снижает риск галлюцинаций.

Разработанный пайплайн обработки заданий и отчётов включает следующие модули:

1. Модуль предварительной обработки. Исходные файлы (задание и отчет) конвертируются в единый markdown-формат с помощью OCR-модели, что позволяет сохранить структуру документа и преобразовывать рисунки и диаграммы в текстовые описания, доступные для анализа большой языковой моделью (далее - БЯМ).
2. Агент извлечения требований. Ключевой элемент системы, решающий проблему универсальности. Агент на базе наиболее передовой модели проводит анализ текста задания лабораторной работы и без использования примеров формирует

список критериев для проверки. Выделяются два типа требований: обязательные (универсальные для всех работ) и специфические (уникальные для конкретной задачи)

3. Агент проверки структуры. Легковесная модель проверяет наличие обязательных разделов отчета, используя сравнение заголовков из списка критериев и автособираемого оглавления работы студента, что допускает вариативность формулировок.
4. Агент проверки содержательных требований. В данном случае модель последовательно проверяет выполнение специфических требований в каждом разделе отчета. Для обработки объемных документов реализован дополнительный механизм сегментации текста: если раздел превышает контекстное окно, перед проверкой вспомогательный подагент разбивает его на логически независимые части.
5. Агент обратной связи. Агрегирует результаты проверок и генерирует обратную связь в зависимости от задаваемого фреймворка. Поддерживаются различные шаблоны, что позволяет предоставлять обратную связь в стиле, наиболее подходящем для преподавателя.

Выводы.

Разработанная система была протестирована на реальных отчётах студентов по ряду дисциплин направления 09.04.04 «Программная инженерия», а также на олимпиадных работах. Система продемонстрировала способность корректно формировать список требований из неструктурированного текста задания без какого-либо предварительно обучения, что решило ключевую проблему исследования. Эксперты в предметной области (преподаватели по тестируемым дисциплинам) подтвердили высокое качество работы всех элементов системы. Предложенная система может быть в дальнейшем усовершенствована за счёт добавления агентов по проверке других аспектов отчёта, дообучения модели по генерации описаний рисунков и совершенствования инструкций к БЯМ для повышения универсальности решения.

Литература

1. UNESCO survey: Two-thirds of higher education institutions have or are developing guidance on AI use [Электронный ресурс]. — 2025. — Режим доступа: <https://www.unesco.org/en/articles/unesco-survey-two-thirds-higher-education-institutions-have-or-are-developing-guidance-ai-use> (дата обращения: 24.02..2026)
2. Emirtekin E. Large language model-powered automated assessment: a systematic review //Applied Sciences. – 2025. – Т. 15. – №. 10. – С. 5683.