

АВТОМАТИЗАЦИЯ РЕЦЕНЗИРОВАНИЯ СТУДЕНЧЕСКИХ РАБОТ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Казиев И. Р¹, Аминов Н. С.²
Научный руководитель – Аминов Н. С.²
Университет ИТМО
ivkaziev@yandex.ru

Работа выполнена в рамках темы НИР №625133 «Методы, модели и архитектуры интеллектуальных сервисов и приложений».

Введение

Рост академической нагрузки и увеличение потоков студентов приводят к дефициту времени у преподавателей на качественную проверку письменных работ (отчетов по лабораторным и практическим занятиям). По данным ВШЭ за 2023 год на одного преподавателя в среднем приходится 5,8 дисциплин, также значительная часть времени (до 30–40 %) уходит на первичный технический контроль: проверку структуры документа, оформления по стандартам и наличия обязательных разделов [1]. Это снижает время, доступное для глубокого анализа содержания и менторства. Исследования последних лет показывают, что большие языковые модели способны ускорить и улучшить процесс аннотирования и рецензирования студенческих работ [3–4], в связи с чем целью работы стала разработка архитектуры информационной системы проверки студенческих работ с использованием интеллектуальных технологий.

Основная часть

В работе предлагается архитектура интеллектуального ассистента проверки работ, относящаяся к третьему поколению систем. Целью разработки является сокращение нагрузки преподавателя за счет автоматизации технического контроля и генерации черновика рецензии. Ключевой особенностью предложения является интеграция с существующей LMS на базе Moodle, что позволяет отказаться от создания отдельного интерфейса и обеспечить бесшовное внедрение в образовательные структуры. Система реализуется в виде плагина к Moodle, взаимодействующего с отдельным сервисом проверки, развернутым в контуре локально для соблюдения требований Федерального закона № 152-ФЗ о персональных данных. Для обеспечения асинхронности и обработки пиковых нагрузок применяются очереди задач и кэширование. Основой семантического анализа служат открытые большие языковые модели, запускаемые локально. Для повышения качества генерации отзыва используется подход RAG (Retrieval-Augmented Generation): запрос модели обогащается релевантными фрагментами из векторной базы данных, содержащей эталонные работы, методические материалы и требования ГОСТ. Дополнительно система включает детерминированные валидаторы для формальной проверки параметров документа (поля, шрифты, структура).

Анализ современной литературы [2–5] выявил ряд проблем, характерных для автоматической проверки работ на основе LLM. Среди основных: склонность моделей к «галлюцинациям», неполнота или избыточность обратной связи, а также проблема «холодного старта», когда для обучения модели требуется огромный массив размеченных данных. Предлагаемая архитектура решает эти проблемы комплексно: использование RAG минимизирует галлюцинации, опираясь на проверенные источники; применение техник Self-Consistency и саморефлексии повышает полноту и релевантность фидбека; а возможность ручного уточнения критериев преподавателем в процессе работы позволяет адаптировать систему без необходимости переобучения.

Ключевым компонентом системы является ядро проверки, архитектура которого вдохновлена современными исследованиями в области автоматизированного оценивания [6].

1. Модуль первичной оценки (GRADER): Получает текст работы, рубрики с весами критериев и набор инструкций (промт), построенный по методике Chain-of-Thought. Модуль идентифицирует в работе смысловые секции, оценивает каждую рубрику и выставляет предварительный балл с обоснованием.
2. Модуль саморефлексии (REFLECTOR): Выполняет критический анализ результата первого этапа. Сравнивая работу с эталонными ответами из векторной базы данных, он выявляет потенциальные ошибки оценивания, пропущенные критерии или несоответствия в логике.
3. Модуль коррекции (REFINER): На основе замечаний, сгенерированных REFLECTOR, формирует итоговый сбалансированный отзыв и финальную оценку. Такой многоступенчатый подход, включающий саморефлексию, повышает согласованность (consistency) оценок и снижает влияние «галлюцинаций» на 10–15% согласно экспериментальным данным. Процесс проверки завершается формированием проекта отзыва, который поступает преподавателю для финальной верификации и утверждения. Такой подход сохраняет ответственность за оценивание за человеком, что снимает этические риски, связанные с работой нейросетей.

Выводы

Предложенное решение позволяет сократить время на первичную проверку работ и ознакомление с их недостатками за счет объединения формальной валидации и интеллектуального анализа содержания. Интеграция с Moodle обеспечивает удобство пользователей и синхронизацию данных с электронным журналом. Использование локальных LLM гарантирует сохранность данных в контуре университета. Планируется реализация плагина Moodle и сервиса проверки с последующим внедрением и апробацией в учебном процессе Университета ИТМО.

Литература

1. Мониторинг экономики образования. НИУ ВШЭ [Электронный ресурс]. – Режим доступа: https://www.hse.ru/data/2023/02/28/2032719144/ib_4%2846%29_2023.pdf
2. Применение больших языковых моделей в образовательном процессе [Электронный ресурс]. – Режим доступа: <https://vestnik.volbi.ru/upload/numbers/368/article-368-4150.pdf>
3. Использование больших языковых моделей для оптимизации процессов обучения в высших учебных заведениях [Электронный ресурс]. – Режим доступа: https://libeldoc.bsuir.by/bits...itov_Using_large_language.pdf
4. Большие языковые модели при решении педагогических задач [Электронный ресурс]. – Режим доступа: <https://vestnik-muiv.ru/upload/iblock/349/yflbyodpr0hrn0alg4pendbaeip7zows.pdf>
5. LLM-generated Feedback in Real Classes and Beyond: Perspectives from Students and Instructors [Электронный ресурс]. – Режим доступа: <https://educationaldatamining.org/edm2024/proceedings/2024.EDM-posters.99/>
6. Advances in Auto-Grading with Large Language Models: A Cross-Disciplinary Survey [Электронный ресурс]. – Режим доступа: <https://aclanthology.org/2025.bea-1.35.pdf>