

ИССЛЕДОВАНИЕ ПРИНЦИПОВ РАЗРАБОТКИ МОДЕЛИ ЛОКАЛЬНОЙ АВТОМАТИЗИРОВАННОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ РЕЦЕНЗИРОВАНИЯ И ОЦЕНИВАНИЯ СТУДЕНЧЕСКИХ РАБОТ НА БАЗЕ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ В РАМКАХ УЧЕБНОГО ПРОЦЕССА ВУЗА



А. К. Красовский, студент
E-mail: andrew.krasovsky39@gmail.com
ФГБОУ ВО «Калининградский государственный
технический университет»

Э. А. Назин, студент
E-mail: nazin.kd131@gmail.com
ФГБОУ ВО «Калининградский государственный
технический университет»

А. Н. Ильяшов, старший преподаватель
E-mail: andrej.ilyashov@klgtu.ru
ФГБОУ ВО «Калининградский государственный
технический университет»

В работе исследуются принципы построения локальной автоматизированной аналитической системы рецензирования и оценивания письменных работ обучающихся высших учебных заведений на основе нейросетевых моделей обработки естественного языка. Обоснована актуальность задачи в условиях растущего объема самостоятельных работ студентов и невозможности применения облачных решений по соображениям конфиденциальности персональных данных. Описана многоуровневая архитектура разработанной системы АСПСР, объединяющая клиентский интерфейс на React 18 / TypeScript / Tailwind CSS, серверный слой на Electron и Node.js, а также локальное AI-ядро на базе llama.cpp и сервера llama-server, исполняющее языковую модель Qwen2.5 7B Q4_K_M. Сформулированы принципы построения системы (локальности, прозрачности, многокритериальности и поддержки преподавателя) и приведены критерии автоматизированного оценивания. Описана методология работы системы: классификация типа работы по совокупности признаков, выбор соответствующей рубрики, формирование промпта и генерация рецензии. Приведены характеристики реализованного прототипа: около 8 360 строк кода, поддержка 12 и более типов работ, среднее время рецензирования одной работы 2–5 мин на типовой пользовательской конфигурации с 8 ГБ оперативной памяти.

Ключевые слова: автоматизированное рецензирование, оценивание студенческих работ, нейросетевые модели, обработка естественного языка, локальные языковые модели, llama.cpp, Qwen2.5, образовательные технологии, контекстно-адаптивное рецензирование, конфиденциальность данных.

ВВЕДЕНИЕ

Современный учебный процесс в высших учебных заведениях характеризуется существенным объемом письменных работ обучающихся: эссе, рефераты, отчеты по лабораторным и практическим занятиям, курсовые проекты и выпускные квалификационные работы. Проверка и рецензирование таких материалов выполняется преподавателями вручную и занимает значительные временные ресурсы. По оценкам исследователей, до 30 % учебной нагрузки преподавателей направлено на проверку письменных заданий, что снижает время,

доступное для непосредственной педагогической работы со студентами и развернутой обратной связи.

Развитие нейросетевых моделей обработки естественного языка, в первую очередь архитектур семейства трансформер [1], создало предпосылки для частичной автоматизации этой деятельности. Современные крупные языковые модели способны выполнять задачи анализа структуры текста, выявления орфографических и стилистических ошибок, оценки соответствия содержания заявленной теме, формирования содержательных замечаний [2, 3]. Однако применение облачных AI-сервисов в учебном процессе сопряжено с рядом существенных ограничений. Передача персональных данных обучающихся и результатов их интеллектуальной деятельности на серверы внешних поставщиков несовместима с требованиями законодательства Российской Федерации в части защиты персональных данных. Стоимость лицензий на корпоративное использование облачных моделей существенна и не всегда соответствует бюджетным возможностям учебных заведений. Существующие специализированные инструменты – системы проверки оригинальности «Антиплагиат» и Turnitin – выполняют лишь поиск заимствований и не осуществляют содержательного анализа работы.

Таким образом, в закрытом контуре учебного заведения отсутствует автономный инструмент для содержательного анализа, рецензирования и формирования персональных рекомендаций обучающимся. Изложенные обстоятельства определяют актуальность разработки локальной автоматизированной системы рецензирования и оценивания студенческих работ (далее – АСПСР), функционирующей на инфраструктуре учебного заведения и использующей нейросетевые модели, развернутые на стороне образовательной организации. Научная новизна работы состоит в формулировке принципов построения такой системы и в проектировании ее архитектуры применительно к специфике учебного процесса российского технического вуза, а также в реализации механизма контекстно-адаптивного рецензирования, при котором набор критериев оценки динамически подбирается на основе автоматически определяемого типа работы.

ОБЪЕКТ ИССЛЕДОВАНИЯ

Объектом исследования является процесс рецензирования и оценивания письменных работ обучающихся в рамках учебного процесса высшего учебного заведения. Предмет исследования – принципы и архитектурные решения построения локальной автоматизированной аналитической системы, выполняющей функции предварительного рецензирования и формирования рекомендаций для преподавателя на основе нейросетевых моделей обработки естественного языка.

В рамках работы рассматриваются работы студентов технических и гуманитарных направлений подготовки: рефераты, эссе, отчеты по лабораторным и практическим занятиям, курсовые проекты, в том числе содержащие фрагменты программного кода и математические формулы. Из рассмотрения исключены работы, требующие оценки экспериментальной части или практической реализации (запуск программного кода, чертежи, физические образцы), – для них требуются специализированные методы оценивания, не входящие в предмет настоящего исследования.

ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Цель работы – разработать локальную автоматизированную аналитическую систему рецензирования и оценивания студенческих работ на базе нейросетевых моделей, ориентированную на применение в учебном процессе технического вуза без передачи данных в сторонние сервисы и формирующую содержательные рекомендации по улучшению работы.

Для достижения поставленной цели сформулированы следующие задачи: 1) провести анализ существующих подходов к автоматизированному оцениванию текстовых работ и определить применимость их методов в контексте учебного процесса российского вуза; 2) сформулировать принципы построения локальной системы рецензирования, учитывающие требования конфиденциальности и поддержки преподавателя; 3) реализовать загрузку и пар-

синг входных файлов в форматах PDF, DOCX, TXT и MD; 4) реализовать механизм автоматического определения типа работы и подбора специализированной рубрики оценки; 5) реализовать формирование промпта и генерацию рецензии локальной языковой моделью; 6) реализовать формирование персонального плана рекомендаций по доработке; 7) обеспечить полную прозрачность процесса через журналирование каждого решения системы; 8) определить условия экспериментальной апробации разработанной модели в учебном процессе.

МЕТОДЫ ИССЛЕДОВАНИЯ

Методологическую основу исследования составляют методы системного анализа, объектно-ориентированного проектирования программных систем, сравнительного анализа существующих программных решений, а также методы машинного обучения и обработки естественного языка применительно к задачам анализа учебных текстов. Для оценки производительности прототипа применялись методы экспериментального измерения времени отклика и потребления вычислительных ресурсов.

Сравнительный анализ выполнен в отношении трех классов существующих решений: облачных LLM-сервисов общего назначения (ChatGPT, Claude, Gemini), систем проверки оригинальности («Антиплагиат», Turnitin) и систем автоматизированного оценивания эссе класса Automated Essay Scoring [4]. Установлено, что ни одна из перечисленных систем не сочетает одновременно полную локальность развертывания, содержательное контекстно-зависимое рецензирование, формирование персональных рекомендаций и адаптацию под тип работы и учебный контекст вуза.

В качестве технологической основы выбран следующий стек: для клиентского интерфейса – библиотека React 18 с языком TypeScript и системой стилей Tailwind CSS, обеспечивающая интерфейс преподавателя и администратора системы; для серверного слоя – фреймворк Electron с прослойкой на Node.js, обеспечивающей управление проверками, маршрутизацию запросов, журналирование и связь с локальными сервисами; для AI-ядра – открытая библиотека инференса трансформерных моделей llama.cpp с надстройкой llama-server, обеспечивающая локальный запуск моделей, обработку текста и формирование ответов [5]. Для развертывания во внутреннем контуре учебного заведения предусмотрена контейнеризация всех компонентов средствами Docker; компоненты «AI-сервис анализа», «Сервис маршрутизации запросов», «Журналирование и аудит», «Управление доступом» собираются в отдельные контейнеры.

В качестве рекомендуемой локальной языковой модели выбрана модель Qwen2.5 7B в формате GGUF с квантованием Q4_K_M [6]. Выбор обоснован сочетанием следующих факторов: модель обладает достаточным качеством обработки русскоязычных текстов и способна корректно работать с фрагментами программного кода; квантование Q4_K_M обеспечивает компромисс между качеством инференса и требованиями к ресурсам – модель занимает около 4 ГБ дисковой памяти и около 6 ГБ оперативной памяти при выполнении, что укладывается в ограничения типовой пользовательской конфигурации с 8 ГБ оперативной памяти. Принципиально архитектура поддерживает замену модели на любую совместимую с GGUF-форматом без изменения остальных компонентов системы.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В ходе исследования сформулированы четыре принципа построения локальной автоматизированной системы рецензирования и оценивания студенческих работ. Принцип локальности предусматривает, что вся обработка данных, включая работу нейросетевой модели, выполняется на инфраструктуре учебного заведения и не предполагает передачи материалов в сторонние сервисы. Принцип прозрачности требует, чтобы каждое решение системы сопровождалось обоснованием и журнальной записью с указанием конкретных фрагментов текста и критериев, на основании которых оно получено. Принцип многокритериальности означает, что итоговая оценка формируется как агрегированная оценка по нескольким независимым критериям, отражающим различные аспекты качества работы. Принцип поддержки

преподавателя устанавливает, что система формирует предварительную рецензию и предложения по доработке, а итоговое академическое решение остается за преподавателем, что соответствует принятым в Российской Федерации подходам к применению искусственного интеллекта в образовательном процессе [7].

На основе сформулированных принципов разработан программный комплекс АСПСР версии 1.2. Объем программного кода составил около 8 360 строк на языке TypeScript, разделенных на 30 файлов; реализованы 8 сервисов бизнес-логики и 16 интерфейсных компонентов. Установочный пакет (без учета файлов модели) занимает около 15 МБ. Поддерживается операционная система Windows 10/11; в качестве установщика используется NSIS, обеспечивающий простое разворачивание на рабочих станциях преподавателей и в учебных аудиториях. Среднее время формирования рецензии для одной работы по пяти критериям составляет 2–5 мин на типовой пользовательской конфигурации.

Пользовательский интерфейс прототипа Smart Grader AI представлен на рисунке 1; он объединяет загрузку работ, выбор модели, рубрику, агентов-экспертов, статистику и историю проверок.

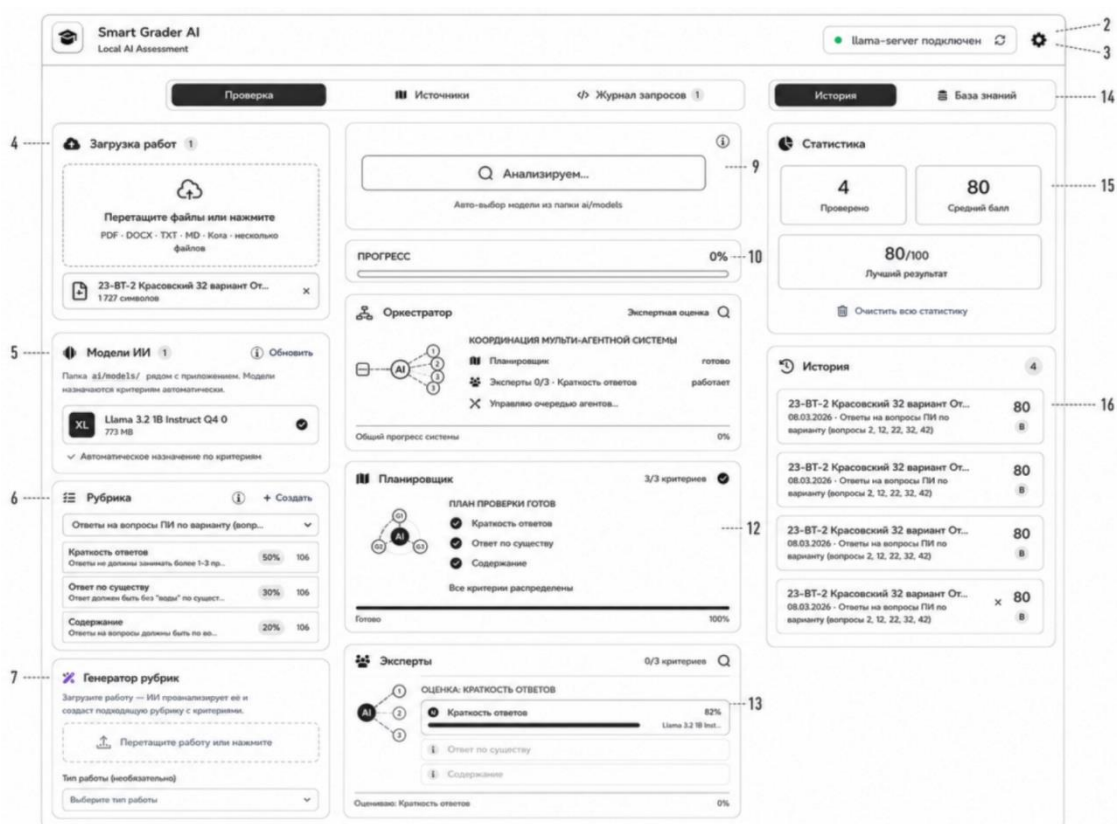


Рисунок 1 – Схема интерфейса АСПСР: 2 – подключение llama-server; 3 – настройки; 4 – загрузка работ; 5 – ИИ-модели; 6 – рубрика; 7 – генератор рубрик; 9 – зона анализа; 10 – прогресс проверки; 12 – планировщик; 13 – эксперты; 14 – история/база знаний; 15 – статистика; 16 – история проверок

Методология работы системы построена в виде последовательной цепочки преобразований входной работы в итоговую рецензию. Соответствующая схема процесса приведена на рисунке 2.

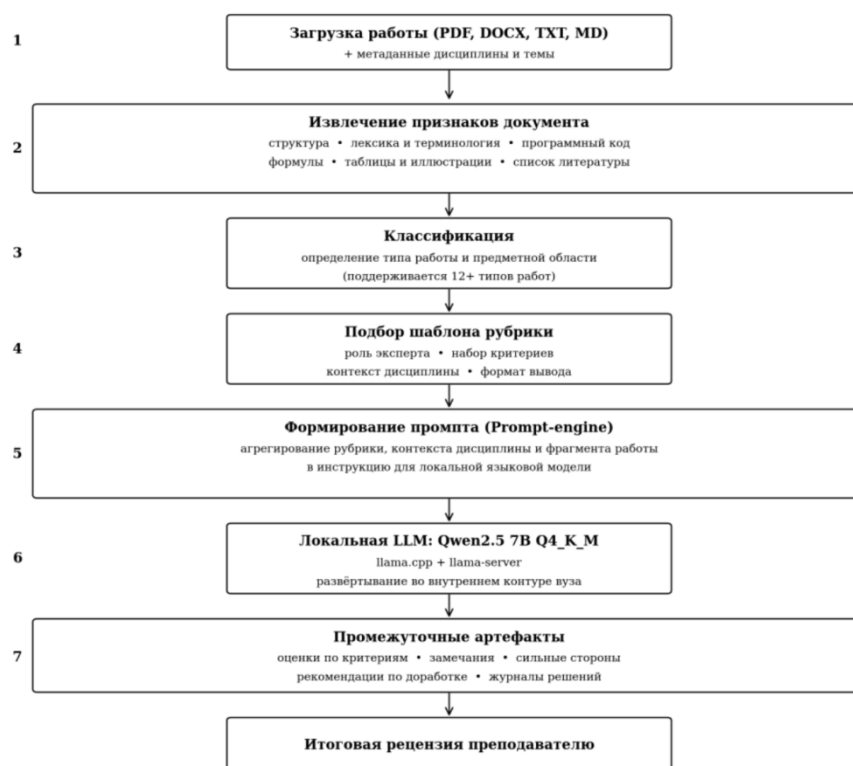


Рисунок 2 – Методология работы системы АСПСР

На этапе извлечения признаков система анализирует структуру документа (заголовки, разделы, оглавление), лексику и терминологию, наличие фрагментов программного кода, математических формул, таблиц, иллюстраций и списка использованных источников. На основе совокупности признаков выполняется классификация: система относит работу к одному из поддерживаемых типов (поддерживается более 12 типов: реферат, эссе, отчет по лабораторной работе, курсовой проект, выпускная квалификационная работа и другие) и определяет ее предметную область. Этот шаг обеспечивает реализацию принципа контекстно-адаптивного рецензирования: для разных типов работ применяются разные критерии оценки. Например, для курсовой работы по программированию анализируются корректность технических решений, структура кода, полнота реализации, качество пояснений и аргументация выбора подходов; для эссе по гуманитарной дисциплине – логика изложения, работа с источниками, аргументация, глубина интерпретации, авторская позиция.

На этапе формирования промпта система собирает инструкцию для языковой модели, объединяющую: роль эксперта, соответствующего предметной области работы; набор критериев оценки, выбранных в соответствии с типом работы; контекст дисциплины; формат итогового ответа. Такой подход (формирование промпта по шаблону рубрики) представляет собой реализацию техники Prompt Engineering применительно к задаче содержательного рецензирования и обеспечивает воспроизводимость результата при повторном анализе одной и той же работы.

На этапе генерации рецензии локальная языковая модель формирует промежуточные артефакты: оценки по критериям, замечания, перечень сильных сторон работы, рекомендации по доработке. Промежуточные артефакты сохраняются в журнале решений системы – это обеспечивает прозрачность работы и возможность последующего аудита. На завершающем этапе артефакты собираются в структурированный отчет, который передается преподавателю через интерфейс приложения. Преподаватель имеет возможность согласиться с автоматизированной оценкой, скорректировать ее или отклонить полностью; финальное академическое решение принимается преподавателем.

Перечень типовых критериев оценивания, реализованных в системе, представлен в таблице 1. Критерии разделены на две группы: формальные критерии оцениваются автома-

тизировано с применением традиционных методов обработки текста; содержательные критерии требуют семантического анализа и обрабатываются языковой моделью. Итоговая агрегированная оценка формируется как взвешенная сумма оценок по отдельным критериям, веса которых задаются в настройках рубрики и могут корректироваться преподавателем.

Таблица 1 – Критерии автоматизированного оценивания студенческих работ

№	Критерий	Содержание оценивания	Группа
1	Структурное соответствие	Наличие требуемых разделов работы и их расположение в установленном порядке	Формальная
2	Соответствие объему	Соответствие объема работы установленным требованиям по количеству страниц	Формальная
3	Орфография и пунктуация	Количество и характер выявленных орфографических и пунктуационных ошибок	Формальная
4	Стиль изложения	Соответствие стиля изложения научно-техническому стилю	Содержательная
5	Соответствие теме	Степень раскрытия заявленной темы и соответствие содержания формулировке темы	Содержательная
6	Логичность изложения	Связность изложения, наличие логических переходов между разделами	Содержательная
7	Аргументированность	Наличие обоснований приводимых утверждений и опоры на источники	Содержательная
8	Качество источников	Количество, актуальность и научный характер использованных источников	Содержательная

Сравнительная характеристика разработанной системы относительно существующих решений приведена в таблице 2. Сравнение проводилось по шести критериям, отражающим ключевые требования к системе автоматизированного рецензирования в условиях учебного процесса.

Таблица 2 – Сравнение АСПСР с существующими классами решений

Критерий	Облачные LLM-сервисы	Антиплагиат-системы	АСПСР
Конфиденциальность данных	Данные передаются на внешние серверы	Требуется отправка в облако	Полностью локально
Глубина содержательного анализа	Ограниченная, без учета учебного контекста	Только поиск заимствований	Контекстно-зависимый анализ
Персональные рекомендации	Общие, не привязаны к критериям	Отсутствуют	Детальные, по разделам
Адаптация под тип работы	Слабая, универсальные ответы	Ограниченная, только текст	Полная (12+ типов)
Развертывание во внутреннем контуре вуза	Нет	Нет (обычно облако)	Да
Прозрачность и аудит решений	«Черный ящик», нет журналов	Ограниченная (только отчет о совпадениях)	Полное журналирование

Из приведенных данных следует, что разработанное решение совмещает свойства, не реализованные одновременно ни в одном из существующих классов решений: полностью локальное развертывание, содержательный контекстно-зависимый анализ, формирование персональных рекомендаций и поддержка большого числа типов работ. Это подтверждает практическую и научную значимость предложенного подхода.

Применение системы предполагает следующий сценарий работы. Обучающийся загружает выполненную работу в систему через интерфейс приложения. Система автоматически определяет тип работы и предметную область, подбирает соответствующую рубрику и формирует структурированную рецензию, включающую оценки по каждому критерию, обоснование оценок, перечень сильных сторон и слабых сторон работы, а также рекомендации по доработке (например, «Усилить тезис во „Введении“, добавив актуальность проблемы и четкую цель исследования»; «Расширить список литературы, добавив минимум три свежие научные статьи по теме за последние пять лет»). Обучающийся получает возможность ознакомиться с рецензией и при необходимости доработать работу до ее представления преподавателю. Преподаватель получает доступ к работе и сформированной рецензии и осуществляет окончательное оценивание.

Экспериментальная апробация модели предполагается в условиях учебного процесса по дисциплинам кафедры информационных и цифровых технологий с участием преподавателей в качестве экспертов. Критерием успешности апробации является степень согласованности оценок, формируемых системой, с оценками преподавателей-экспертов: показатель согласованности рассчитывается как доля совпадений автоматизированной оценки с экспертной с допустимым отклонением в пределах одного балла по пятибалльной шкале. Дополнительно оценивается субъективное восприятие качества рецензий студентами и преподавателями по разработанной анкете.

ЗАКЛЮЧЕНИЕ

В работе исследованы принципы построения локальной автоматизированной аналитической системы рецензирования и оценивания письменных работ обучающихся высших учебных заведений на основе нейросетевых моделей и реализован прототип системы АСПСР версии 1.2. Обоснована необходимость локального развертывания подобной системы в условиях, когда облачные решения неприменимы по соображениям конфиденциальности персональных данных. Сформулированы четыре принципа построения системы: локальности, прозрачности, многокритериальности и поддержки преподавателя, и разработана многоуровневая архитектура, реализующая эти принципы на стеке React 18 / TypeScript / Tailwind CSS, Electron / Node.js, llama.cpp / llama-server с моделью Qwen2.5 7B Q4_K_M.

Научная значимость результатов состоит в систематизации принципов и архитектурных решений, применимых к задачам автоматизированного содержательного рецензирования учебных работ в условиях ограничений на передачу данных, и в обосновании механизма контекстно-адаптивного рецензирования с автоматическим подбором рубрики оценки по типу работы. Практическая значимость подтверждается работающим прототипом объемом около 8 360 строк кода, поддерживающим более 12 типов работ и обеспечивающим формирование рецензии за 2–5 мин на типовой пользовательской конфигурации, и возможностью использования результатов исследования при разработке программного комплекса, ориентированного на интеграцию в учебный процесс конкретного учебного заведения.

Дальнейшие направления исследования включают экспериментальную апробацию предложенной модели на реальных студенческих работах с привлечением преподавателей-экспертов, расширение поддержки на операционные системы Linux и macOS, интеграцию с системами управления обучением (Moodle, Canvas), реализацию пакетной проверки папок с работами, а также исследование возможности дообучения локальной нейросетевой модели на корпусах учебных текстов конкретного учебного заведения.

СПИСОК ЛИТЕРАТУРЫ

1. Vaswani, A. Attention is all you need / A. Vaswani, N. Shazeer, N. Parmar [et al.] // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998–6008.
2. Brown, T. B. Language models are few-shot learners / T. B. Brown, B. Mann, N. Ryder [et al.] // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 1877–1901.
3. Mizumoto, A. Exploring the potential of using an AI language model for automated essay scoring / A. Mizumoto, M. Eguchi // Research Methods in Applied Linguistics. – 2023. – Vol. 2, № 2. – Article 100050.
4. Ke, Z. Automated essay scoring: A survey of the state of the art / Z. Ke, V. Ng // Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI). – 2019. – P. 6300–6308.
5. llama.cpp [Электронный ресурс]: репозиторий проекта на GitHub. – URL: <https://github.com/ggml-org/llama.cpp> (дата обращения: 01.05.2026). – Текст: электронный.
6. Yang, A. Qwen2.5 Technical Report / A. Yang, B. Yang, B. Zhang [et al.] // arXiv preprint arXiv:2412.15115. – 2024.
7. ГОСТ Р 59276–2020. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения. – Москва: Стандартинформ, 2020.

STUDY OF THE PRINCIPLES OF DEVELOPING A LOCAL AUTOMATED ANALYTICAL SYSTEM FOR REVIEWING AND GRADING STUDENT WORKS BASED ON NEURAL NETWORK MODELS WITHIN THE EDUCATIONAL PROCESS OF A UNIVERSITY

A. K. Krasovskii, student
E-mail: andrew.krasovsky39@gmail.com
Kaliningrad State Technical University

E. A. Nazin, student
E-mail: nazin.kd131@gmail.com
Kaliningrad State Technical University

The paper presents the development of a local automated analytical system for reviewing and grading written works of higher education students based on neural network natural language processing models. The relevance of the task is determined by the growing volume of independent works performed by students and by the impossibility of applying cloud-based solutions due to the confidentiality requirements imposed on personal data and on the results of students' intellectual activity. A multi-level architecture of the system named ASPSR (version 1.2) is described, comprising a client interface based on React 18, TypeScript and Tailwind CSS, a server layer based on Electron and Node.js, and a local AI core based on llama.cpp with the llama-server wrapper running the Qwen2.5 7B Q4_K_M language model. Four principles of system design have been formulated: locality, transparency, multi-criteria assessment and support for the teacher. The methodology of system operation has been described, including feature extraction from the input document, automatic classification of the work type and subject domain, selection of the corresponding evaluation rubric, prompt formation through the prompt engine, and generation of the structured review by the local language model. A set of grading criteria covering both formal and content-related aspects of student works has been proposed. A comparison with existing classes of solutions (cloud LLM services, plagiarism detection systems) is provided. The implemented prototype consists of approximately 8 360 lines of TypeScript code, supports more than 12 types of works, and forms a review of one work in 2 to 5 minutes on a standard user configuration with 8 GB RAM.

Keywords: *automated review, student works grading, neural network models, natural language processing, local language models, llama.cpp, Qwen2.5, educational technologies, context-adaptive review, data confidentiality.*