

ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ЧАТ-БОТА ДЛЯ ГЕНЕРАЦИИ РЕКОМЕНДАЦИЙ ПО ИНКЛЮЗИВНОМУ ОБРАЗОВАНИЮ С ПРИМЕНЕНИЕМ RAG-ТЕХНОЛОГИЙ



М.М. Тарасов

*Магистрант Высшей школы электроники и компьютерных наук
ФГАОУ ВО «ЮУрГУ (НИУ)»
tosik4302@mail.ru*



Л.Ю. Костылева

*Доцент кафедры «Информационные системы и технологии» Высшей школы электроники и компьютерных наук ФГАОУ ВО «ЮУрГУ (НИУ)», кандидат технических наук
kostylevali@susu.ru*

М.М. Тарасов

Окончил Южно-Уральский государственный университет (национальный исследовательский университет). Область научных интересов связана с разработкой информационных систем, методов и алгоритмов машинного обучения.

Л.Ю. Костылева

Окончила Южно-Уральский государственный университет (национальный исследовательский университет). Область научных интересов связана с исследованием проблем управления в организационных системах, методов и алгоритмов совершенствования управленческой деятельности.

Аннотация. В статье рассматривается концепция и архитектура разрабатываемого чат-бота, предназначенного для помощи преподавателям высших учебных заведений в организации образовательного процесса для студентов с ограниченными возможностями здоровья (ОВЗ). Предлагаемое решение базируется на современной RAG-технологии (Retrieval-Augmented Generation), которая позволяет интегрировать большие языковые модели с актуальной нормативно-методической базой. В качестве источника знаний выступают официальные методические пособия Министерства науки и высшего образования РФ. Описываются функциональные модули системы, обосновывается выбор технологического стека и приводятся ожидаемые результаты внедрения.

Ключевые слова: RAG-технология, чат-бот, инклюзивное образование, студенты с ОВЗ, искусственный интеллект в образовании, рекомендательная система.

Введение. Современная система высшего образования в Российской Федерации характеризуется устойчивым развитием инклюзивных практик и расширением доступности образовательных услуг для различных категорий граждан. В последние годы наблюдается рост числа студентов с ограниченными возможностями здоровья (ОВЗ), поступающих в вузы. Это связано как с совершенствованием нормативно-правовой базы, так и с формированием более доступной образовательной среды [1].

Ключевым органом, отвечающим за нормативное регулирование в этой области, выступает Министерство науки и высшего образования РФ. Именно оно разрабатывает подзаконные акты и инструктивные письма, регламентирующие процесс обучения студентов с инвалидностью и ограниченными возможностями здоровья. Базовые гарантии права на образование, а также описание механизмов создания адаптированных условий закреплены в Федеральном законе «Об образовании в Российской Федерации» [2].

Современное инклюзивное образование в вузах нельзя сводить только к установке пандусов или специального оборудования. Ключевым фактором становится готовность педагогов взаимодействовать с обучающимися, имеющими разные нозологии. Это требует от преподавателя понимания специфики усвоения материала, умения корректировать содержание курсов, подбирать адекватные формы контроля и выстраивать коммуникацию с учетом индивидуальных особенностей студента. Реализация таких задач невозможна без дополнительных психолого-педагогических знаний и способности быстро перестраивать методику работы.

На первый план выходят такие инструменты, как проектирование персональных траекторий обучения, внедрение дистанционных форматов, корректировка оценочных средств и использование вариативных способов подачи информации. Известны различные успешные способы применения современных цифровых технологий, в том числе чат-ботов и социальных роботов в инклюзивном образовании [3], направленные на персонализированную поддержку студентов с различными образовательными потребностями и ограниченными возможностями здоровья, которые охватывают широкий спектр нозологий, включая нарушения зрения, слуха и другие специфические трудности обучения [4].

Тем не менее, на деле внедрение современных технологий сталкивается с препятствиями в виде высокой загруженности преподавательского состава и запутанности нормативной документации. Действующие регламенты, касающиеся организации учебного процесса для студентов с ОВЗ, отличаются большим объемом и требуют детального изучения. Как следствие, оперативное нахождение ответов на возникающие в конкретной педагогической ситуации вопросы превращается в серьезную проблему [5].

Хотя официальные методические рекомендации существуют, использовать их в ежедневной рутине непросто. Перед преподавателем часто встают частные, но важные вопросы: как скорректировать формат текущего контроля, что должно быть отражено в рабочей программе, какие именно специальные условия обязательны к исполнению. Сбор информации из множества разрозненных источников отнимает много времени и чреват ошибками, связанными с неполнотой или неверной интерпретацией требований [6].

В эпоху цифровизации образования все более востребованными становятся интеллектуальные помощники и системы поддержки принятия решений. Использование чат-ботов для быстрого извлечения информации из загруженной документации представляет собой современный подход к автоматизации работы с документами, который сочетает технологии обработки естественного языка, машинного обучения и информационного поиска. Современные системы обработки естественного языка дают пользователям возможность формулировать запросы в обычной речи и получать точные ответы из текстовых документов без ручного анализа [7]. Алгоритмы на базе трансформеров очень эффективны в работе с текстом – модели BERT показывают высокую точность при классификации документов и при извлечении релевантной информации, уменьшают вероятность человеческих ошибок и заметно ускоряют обработку текстовых данных по сравнению с традиционным анализом [8]. Развитие больших языковых моделей позволяет автоматизировать обработку текстов и генерировать ответы в форме, удобной для человека. Но если применять такие модели в текущем виде без проверки информации по надёжным источникам, это может привести к неправильным или искажённым сведениям.

Эффективной альтернативой является технология дополненной генерации ответов – она объединяет возможности языковой модели с механизмом поиска по авторитетной базе знаний. Благодаря этому итоговый ответ строится на основе фрагментов актуальных официальных документов, что гарантирует его точность и достоверность. Для образовательной практики, где строгое соблюдение нормативов принципиально важно, такой подход представляется наиболее перспективным [9].

В настоящей работе предлагается концепция чат-бота, выполняющего функции интеллектуального ассистента для преподавателей вузов. Бот реализует RAG-архитектуру и обеспечивает диалоговый формат взаимодействия. Пользователь формулирует запрос в свободной форме, после чего система осуществляет поиск релевантных положений в базе нормативных и методических документов и генерирует структурированные рекомендации.

Предлагаемое решение направлено на повышение оперативности принятия педагогических решений, снижение временных затрат на поиск информации и повышение нормативной корректности организации обучения студентов с ОВЗ. Новизна подхода заключается в интеграции RAG-технологии с официальными методическими материалами и ориентации на практическую поддержку преподавателей в режиме реального времени. Разработка подобной системы соответствует современным тенденциям цифровизации высшего образования и отвечает актуальным потребностям инклюзивной образовательной среды.

Постановка задачи и формирование требований к системе. Разработка и создание прототипа чат-бота, способного на основе запроса преподавателя формировать релевантные, нормативно обоснованные рекомендации по адаптации образовательного процесса для студентов с различными нозологиями ОВЗ, обусловлено необходимостью повышения качества инклюзивного образования и снижения временных затрат преподавателей на поиск и интерпретацию регламентирующих документов. Предполагается, что система будет функционировать как интеллектуальный ассистент, обеспечивающий оперативный доступ к структурированной базе знаний, сформированной на основе официальных методических материалов и нормативных документов. Разрабатываемая система не заменяет эксперта в области инклюзии, но может выступать в роли инструмента поддержки принятия решений, повышая обоснованность, корректность и уверенность педагогов в своих действиях.

Для достижения поставленной цели необходимо решить ряд взаимосвязанных задач.

1. Анализ и систематизация источников базы знаний. Нужно провести отбор, структурирование и тематическую классификацию официальных нормативных документов и методических пособий, разработанных Министерством науки и высшего образования

Российской Федерации, а также иных регламентирующих материалов, касающихся организации обучения студентов с ОВЗ. Собранные документы необходимо обработать, выделяя фрагменты, в которых содержатся практико-ориентированные рекомендации. Имеющиеся в вузе локальные нормативные документы и актуальные наработки в области реализации инклюзивного образования также могут быть включены в базу знаний. В ходе систематизации источников следует также запланировать их будущие обновления – продумать способ оценки текущей версии, перечень критериев и атрибутов, позволяющих квалифицировать источник как актуальный либо устаревший, порядок поиска обновлённых версий уже задействованных материалов и правила добавления новых источников в рабочий массив данных.

2. Проектирование архитектуры RAG-системы. Требуется разработать модель архитектуры, включающую модуль предварительной обработки и векторизации текстовых документов с учетом формата файлов, векторную базу данных для хранения эмбедингов и механизм поиска релевантных фрагментов. Архитектура должна обеспечивать масштабируемость, возможность обновления базы знаний и корректную обработку запросов различной сложности.

3. Интеграция генеративной модели. Интеграция должна предусматривать формирование промптов с учетом найденных фрагментов документов, а также настройку параметров генерации для обеспечения точности, формальной корректности, минимизации придуманных ответов, так называемых «галлюцинаций», и стилистической нейтральности ответов. В качестве генеративного ядра системы планируется использование свободно распространяемых генеративных трансформеров (GPT-моделей), подобранных с учетом имеющихся вычислительных возможностей.

4. Разработка пользовательского интерфейса. Взаимодействие пользователя с системой предполагается реализовать через веб-интерфейс или доступный мессенджер. Интерфейс должен обеспечивать удобный ввод запросов на естественном языке, получение структурированных ответов с отображением ссылок на источники, поддержку диалогового формата общения, возможность копирования текста вопросов и ответов, ведение структурированного журнала запросов и ответов с целью последующего анализа диалоговых паттернов.

5. Апробация и оценка эффективности. Планируется проведение пилотного внедрения прототипа в образовательной среде Южно-Уральского государственного университета. В рамках апробации предполагается анализ точности ответов, релевантности рекомендаций, удовлетворенности пользователей и оценка влияния системы на скорость принятия педагогических решений.

Ключевые функциональные требования к системе включают:

– приём и корректную интерпретацию запросов, сформулированных на естественном языке (например, «Какие требования к освещению в аудитории для слабовидящего студента?», «Как адаптировать лабораторную работу для студента с нарушением опорно-двигательного аппарата?»);

– автоматизированный поиск релевантных данных в базе знаний с применением алгоритмов векторного сходства для сопоставления семантики запроса и содержимого базы;

– генерацию связного и структурированного ответа, обладающего практической ценностью, с обязательным указанием источника информации или нормативного основания из документов, находящихся в базе знаний;

– поддержку истории диалога и учёт контекста предыдущих взаимодействий для повышения точности и релевантности рекомендаций;

– обеспечение защиты персональных данных пользователей и устойчивости функционирования системы в условиях повышенной нагрузки.

Дополнительно система должна отвечать требованиям надежности, расширяемости и возможности последующего масштабирования. В перспективе архитектура должна позволять обновлять базу знаний при появлении новых нормативных документов без существенной переработки программной части.

Предлагаемая архитектура решения. Предлагаемая архитектура построена на принципах Retrieval-Augmented Generation (RAG) [9] и включает несколько логически взаимосвязанных модулей, реализующих функциональные требования к системе и обеспечивающих полный цикл обработки запроса – от загрузки нормативных документов до формирования финального ответа преподавателю (рисунок 1).

1. Модуль загрузки и обработки данных. Источником знаний выступают официальные PDF-документы, размещенные на портале Министерства науки и высшего образования Российской Федерации и других официальных сетевых ресурсах – нормативно-правовые акты, методические рекомендации, инструкции, регламенты и иные материалы, касающиеся организации обучения студентов с ОБЗ.

На этапе загрузки документов реализуется автоматизированный конвейер обработки данных, обеспечивающий извлечение текста из PDF-файлов, очистку от служебной информации (колонтитулы, номера страниц, повторяющиеся заголовки), нормализацию текста (удаление лишних пробелов, приведение кодировки), разбиение на логические фрагменты (чанкинг) и извлечение метаданных.

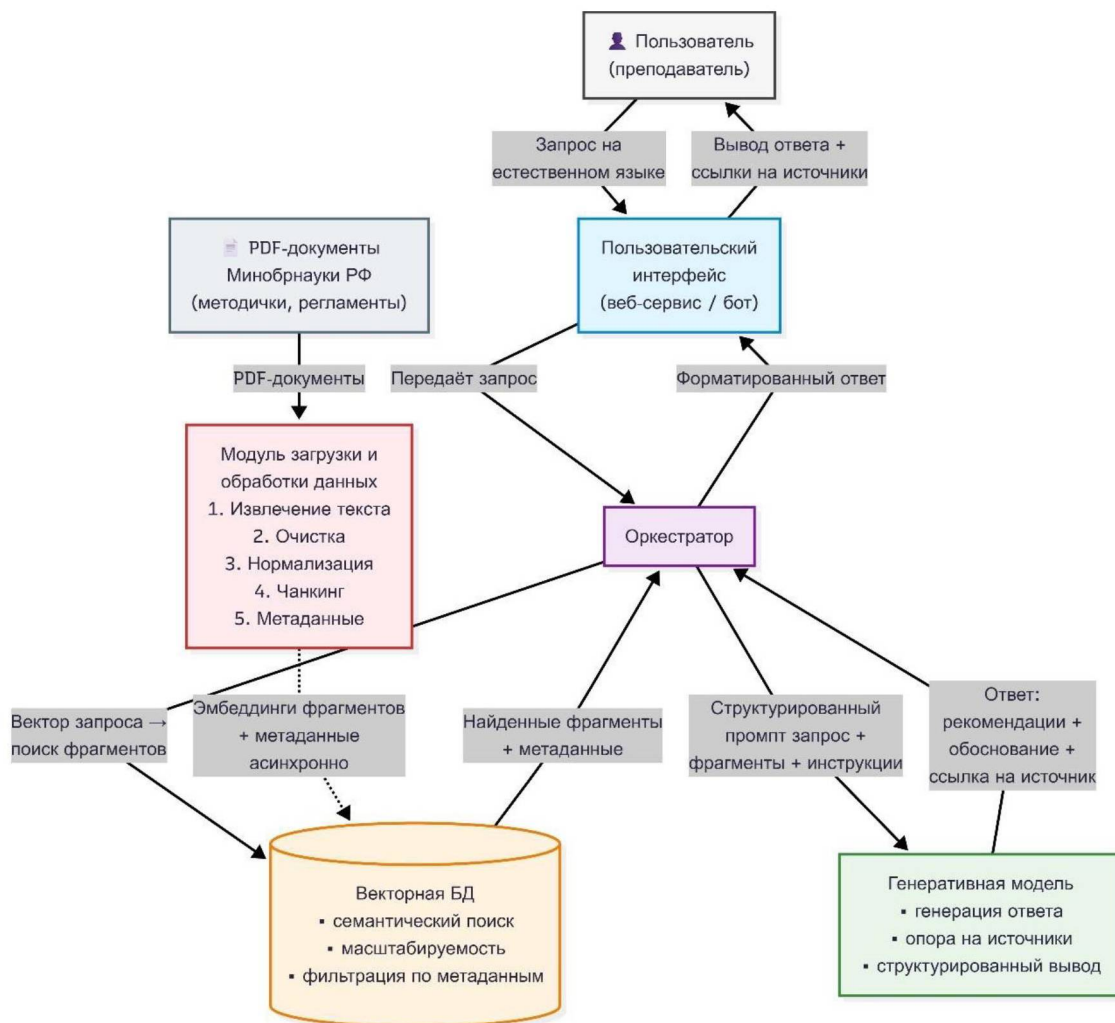


Рисунок 1. Концепция архитектуры чат-бота

Чанкинг осуществляется с учетом смысловой структуры документа, включающей соответствующие разделы, подпункты, абзацы. Размер фрагмента подбирается с учетом контекстного окна применяемой GPT-модели таким образом, чтобы сохранялась семантическая целостность, но при этом обеспечивалась точность поиска. Каждый фрагмент снабжается метаданными (например, название документа, раздел, номер страницы, дата публикации), что в дальнейшем позволяет формировать корректные ссылки на источник.

2. Векторная база данных. После разбиения текстовые фрагменты преобразуются в векторные представления (эмбединги) с использованием специализированной модели. Эмбединг отражает семантическое содержание текста в виде числового вектора фиксированной размерности.

Полученные векторы сохраняются в векторной базе данных (например, ChromaDB, FAISS или Qdrant). Использование векторной БД обеспечивает быстрый семантический поиск по принципу близости, масштабируемость при увеличении объема документов, а также возможность фильтрации по метаданным (тип документа, год и т.д.).

Поиск релевантных фрагментов осуществляется по косинусной мере сходства между вектором пользовательского запроса и векторами, хранящимися в базе.

3. Оркестратор. Представляет собой центральный модуль системы, координирующий взаимодействие между компонентами. Получив запрос от преподавателя через интерфейс чат-бота, он выполняет следующие шаги:

- 1) преобразует текст запроса в вектор с использованием той же модели эмбедингов;
- 2) выполняет поиск в векторной базе данных и извлекает наиболее релевантные фрагменты;
- 3) с помощью специальной модели (ранкера) структурирует извлеченные фрагменты по приоритету: самые релевантные фрагменты первыми подаются в генеративную модель, повышая точность и качество ответа;
- 4) формирует структурированный промпт для генеративной модели, включающий оригинальный запрос пользователя, извлеченные фрагменты нормативных документов и инструкции модели о необходимости опираться исключительно на предоставленные источники.

Таким образом обеспечивается контролируемая генерация ответа и минимизация риска появления недостоверной информации (галлюцинаций).

4. Генеративная модель. Модель получает подготовленный промпт и формирует связный, структурированный ответ, включающий конкретные рекомендации, нормативное обоснование и ссылку на источник (по метаданным фрагмента).

Ключевым принципом является ограничение генерации рамками предоставленных документов. Модель синтезирует ответ только на основе найденных фрагментов, что повышает надежность системы. В качестве генеративного ядра системы планируется использовать доступные GPT-модели, которые могут функционировать во внутреннем контуре организации и имеют удобный API.

5. Пользовательский интерфейс. Может быть реализован в нескольких вариантах – веб-сервиса или бота для доступного мессенджера. Интерфейс обеспечивает прием запросов на естественном языке, вывод структурированных ответов, отображение ссылок на источники и хранение истории диалога для учета контекста.

Описанная архитектура системы ориентирована на модульность, масштабируемость и возможность дальнейшего расширения функциональности.

Ожидаемые результаты и перспективы внедрения. На текущем этапе разработка системы находится в стадии формирования базового прототипа. Реализован чат-бот, обеспечивающий пользовательский интерфейс и обработку входящих сообщений, однако интеграция с полнофункциональным RAG-контуром и генеративной моделью еще не завершена. В настоящее время бот выполняет демонстрационную функцию и служит

платформой для тестирования логики взаимодействия с пользователем. Основные усилия следующего этапа разработки будут направлены на построение векторной базы знаний, настройку алгоритмов семантического поиска и интеграцию с генеративной моделью через API.

Ключевым шагом станет формирование структурированной базы знаний на основе официальных методических документов, их векторизация и размещение в специализированной базе данных. После завершения данного этапа система сможет обеспечивать релевантный поиск фрагментов и формировать ответы, опираясь на нормативные источники. Для оценки эффективности разработанного решения нужно выполнить анализ того, насколько точно извлекаются текстовые данные, грамотно ли выполнено деление документов на логические блоки и правильно ли сопоставляются запросы пользователей с релевантными фрагментами нормативных и методических источников.

Ожидается, что интеграция созданного чат-бота в деятельность Южно-Уральского государственного университета приведет к ряду ощутимых результатов – и в практической плоскости, и в организации учебного процесса. Педагоги будут тратить меньше времени на поиск нужной информации. Вместо того чтобы изучать многостраничные регламенты, преподаватель может просто задать вопрос и сразу получить понятный структурированный ответ с ссылкой на конкретный документ. Фокус сместится с бюрократических процедур на само обучение студентов, это ускорит подготовку к занятиям.

Кроме того, система призвана улучшить качество разработки адаптированных образовательных программ и планирования отдельных занятий с учетом потребностей студентов с ограниченными возможностями здоровья. Благодаря семантическому анализу и опоре на свежие нормативные данные, бот будет предлагать решения, которые учитывают характер заболевания, тип учебного занятия и выбранную форму проверки знаний. Это позволит перейти к более продуманной и методологически выверенной инклюзивной политике в университете.

В психологическом плане ожидается, что появление интеллектуального помощника снизит тревожность и неуверенность преподавателей, поскольку вопросы, касающиеся обучения студентов с ОВЗ, зачастую порождают чувство неопределенности и страх нарушить существующие предписания. Наличие инструмента, предоставляющего обоснованные рекомендации со ссылкой на официальные документы, призвано повысить уверенность педагогов в корректности принимаемых решений.

В перспективе бот может стать единым цифровым окном доступа к актуальной нормативной информации в сфере инклюзивного образования. Централизация источников в рамках одной системы упростит навигацию и обеспечит единообразие интерпретации требований. В дальнейшем это может способствовать формированию единой методической культуры работы со студентами с ОВЗ внутри университета.

Развитие проекта предполагает расширение базы знаний за счет включения локальных нормативных актов университета, регламентов кафедр, положений о текущем и промежуточном контроле, а также доступных лучших практик других образовательных организаций. Возможна интеграция аналитических модулей, позволяющих выявлять наиболее частые запросы преподавателей и формировать рекомендации по совершенствованию внутренних регламентов.

Таким образом, реализация и последующее масштабирование системы имеют потенциал стать значимым элементом цифровой трансформации инклюзивного образования, обеспечивая устойчивую поддержку преподавателей и повышение качества образовательного процесса в целом.

Заключение. В статье представлена концепция интеллектуального чат-бота, предназначенного для поддержки преподавателей высших учебных заведений в вопросах организации инклюзивного образовательного процесса. Предложенное решение

ориентировано на формирование удобного цифрового инструмента, обеспечивающего оперативный доступ к нормативной и методической информации, связанной с обучением студентов с ограниченными возможностями здоровья. В основе архитектуры системы лежит применение технологии RAG, сочетающей семантический поиск по базе знаний с возможностями генеративной модели. Такой подход позволяет формировать ответы, опирающиеся на официальные документы и методические рекомендации, что обеспечивает высокую релевантность, нормативную корректность и практическую применимость выдаваемых рекомендаций. Использование диалогового формата взаимодействия делает систему доступной и удобной в повседневной работе преподавателя. Предполагается, что дальнейшее развитие системы позволит не только повысить оперативность принятия педагогических решений, но и укрепить методическую культуру инклюзивного образования в университете.

Социальная значимость проекта заключается в его направленности на создание более доступной, комфортной и справедливой образовательной среды. Внедрение подобного инструмента может способствовать снижению барьеров в обучении студентов с ОВЗ и повышению качества их образовательного опыта, что соответствует современным приоритетам развития высшего образования.

Список литературы

- [1] Инклюзивная образовательная среда вуза (обзор зарубежных исследований) / С. Т. Кохан, С. В. Власова, П. Ангелова, С. Павлович // Ученые Записки Забайкальского Государственного Университета. – 2024. – Т. 19. – № 2. – С. 41–50.
- [2] Федеральный закон «Об образовании в Российской Федерации» от 29.12.2012 N 273-ФЗ (последняя редакция) \ КонсультантПлюс. – URL: https://www.consultant.ru/document/cons_doc_LAW_140174/ (дата обращения: 29.03.2026). – Текст : электронный.
- [3] Social Robots in Special Education: A Systematic Review / G. A. Papakostas, G. K. Sidiropoulos, C. I. Papadopoulou [et al.] // Electronics. – 2021. – Vol. 10. – Social Robots in Special Education. – № 12. – P. 1398. – DOI: 10.3390/electronics10121398.
- [4] Pérez, J. Q. Rediscovering the use of chatbots in education: A systematic literature review / J. Q. Pérez, T. Daradoumis, J. M. M. Puig // Computer Applications in Engineering Education. – 2020. – Vol. 28. – Rediscovering the use of chatbots in education. – № 6. – P. 1549–1565. – DOI: 10.1002/cae.22326.
- [5] Рубцов, В. В. Инклюзивное образование: актуальный дискурс и перспективы развития / В. В. Рубцов, С. В. Алехина // Психологическая наука и образование. – 2024. – Т. 29. – Инклюзивное образование. – № 5. – С. 4.
- [6] Салаватова, А. М. Нормативно-правовое обеспечение применения информационных технологий в социальной сфере, связанной с реабилитацией людей с инвалидностью / А. М. Салаватова, О. И. Истрофилова, Д. В. Ершов. – ООО «Издательский дом «Среда», 2024. – С. 19–22.
- [7] Conversations with Documents: An Exploration of Document-Centered Assistance / M. ter Hoeve, R. Sim, E. Nouri [и др.]. – Текст : электронный // Proceedings of the 2020 Conference on Human Information Interaction and Retrieval : CHIIR '20. – New York, NY, USA : Association for Computing Machinery, 2020. – С. 43–52. – URL: <https://dl.acm.org/doi/10.1145/3343413.3377971> (дата обращения: 19.03.2026).
- [8] A Comprehensive Analysis of Deep Learning Techniques for Documentation Classification / V. Kumar, A. Dabas, I. Jain, K. Pawar. – Текст : электронный // 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). – Coimbatore, India : IEEE, 2021. – С. 228–235. – URL: <https://ieeexplore.ieee.org/document/9395861/> (дата обращения: 19.03.2026).
- [9] Retrieval-augmented generation for knowledge-intensive NLP tasks / P. Lewis, E. Perez, A. Piktus [et al.] – Текст : электронный // Proceedings of the 34th International Conference on Neural Information Processing Systems : NIPS '20. – Red Hook, NY, USA : Curran Associates Inc., 2020. – URL: <https://dl.acm.org/doi/abs/10.5555/3495724.3496517> (дата обращения 19.03.2026).

Авторский вклад

Тарасов Максим Михайлович – постановка задачи и проведение исследования, разработка требований, концепции и архитектуры системы, алгоритмов обработки данных.

Костьлева Лилия Юрьевна – постановка задачи исследования, анализ источников, общее редактирование и структурирование текста, подготовка к публикации.

DESIGN AND DEVELOPMENT OF A RAG-ENABLED CHATBOT FOR GENERATING RECOMMENDATIONS IN INCLUSIVE EDUCATION

M.M. Tarasov

*Master's student of School of Electronic
Engineering and Computer Science, SUSU*

L. Yu. Kostyleva

*Associate Professor of Department of Information Systems
and Technologies, School of Electronic Engineering and
Computer Science, SUSU,
PhD of Technical Sciences*

Abstract. The article examines the concept and architecture of a developing chatbot designed to assist faculty instructors at higher education institutions in organizing the educational process for students with disabilities (limited health capabilities). The proposed solution is based on modern Retrieval-Augmented Generation (RAG) technology, which enables the integration of large language models with up-to-date regulatory and methodological resources. Official methodological guidelines from the Russian Ministry of Science and Higher Education serve as the knowledge source. The system's functional modules are described, the rationale for the technology stack is justified, and the expected outcomes of implementation are outlined.

Keywords: Retrieval-Augmented Generation (RAG) technology, chatbot, inclusive education, students with disabilities, artificial intelligence in education, recommender system.