

УДК 004.048

## ПРОТОТИП ИИ-АГЕНТА ПО УЧЕБНОМУ КУРСУ КАК ЧАСТЬ УМК

Петросян Анна Суменовна, студент, Базовая кафедра «Аналитика больших данных и методы видеоанализа», Уральский федеральный университет имени первого Президента России Б.Н. Ельцина

Горбатенков Тимофей Владимирович, студент, Базовая кафедра «Аналитика больших данных и методы видеоанализа», Уральский федеральный университет имени первого Президента России Б.Н. Ельцина

Колчанов Константин Викторович, студент, Базовая кафедра «Аналитика больших данных и методы видеоанализа», Уральский федеральный университет имени первого Президента России Б.Н. Ельцина

Мягченко Илья Дмитриевич, студент, Базовая кафедра «Аналитика больших данных и методы видеоанализа», Уральский федеральный университет имени первого Президента России Б.Н. Ельцина

Саиф Муджахед Абдулла Хаель, старший преподаватель Базовой кафедры «Аналитика больших данных и методы видеоанализа», Уральский федеральный университет имени первого Президента России Б.Н. Ельцина

### Аннотация

*В статье рассматривается прототип интеллектуального агента (ИИ-агента), интегрированного в учебно-методический комплекс (УМК) по дисциплине "Практическое предпринимательство". Основная цель исследования – повышение эффективности самостоятельной учебной деятельности студентов и обеспечение гибкого сопровождения образовательного процесса. Разработана архитектура ИИ-агента, предназначенная для адаптивной поддержки обучения и упрощения объяснений студентам при одновременном повышении удобства работы преподавателей.*

**КЛЮЧЕВЫЕ СЛОВА:** интеллектуальный агент, учебно-методический комплекс, адаптивное обучение, персонализация обучения, транскрибация.

## AI-AGENT PROTOTYPE FOR THE EDUCATIONAL COURSE AS PART OF THE EDUCATIONAL AND MULTIPLE-COURSE

Petrosyan Anna Surenovna, Student, Department of Big Data Analytics and Video Analysis Methods, Ural Federal University named after the first President of Russia B.N. Yeltsin

Gorbatenkov Timofey Vladimirovich, Student, Department of Big Data Analytics and Video Analysis Methods, Ural Federal University named after the first President of Russia B.N. Yeltsin

Kolchanov Konstantin Viktorovich, Student, Department of Big Data Analytics and Video Analysis Methods, Ural Federal University named after the first President of Russia B.N. Yeltsin

Myagchenko Ilya Dmitrievich, Student, Department of Big Data Analytics and Video Analysis Methods, Ural Federal University named after the first President of Russia B.N. Yeltsin

Saif Mujahed Abdullah Hayel, Senior Lecturer, Department of Big Data Analytics and Video Analysis Methods, Ural Federal University named after the first President of Russia B.N. Yeltsin

### Abstract

*This article examines a prototype of an intelligent agent (AI agent) integrated into a teaching and learning package (TMP) for the course "Practical Entrepreneurship." The primary goal of the study is to improve the effectiveness of students' independent learning activities and provide flexible support for the educational process. An AI agent architecture has been developed to provide adaptive learning support and simplify explanations to students while simultaneously improving the usability of instructors.*

KEYWORDS: AI agent, teaching and learning package, adaptive learning, personalized learning, transcription

## Введение

Цифровизация высшего образования усилила интерес к персонализированным технологиям обучения, способным поддерживать студентов в самостоятельной работе. Одним из наиболее эффективных направлений считаются интеллектуальные тьюторские системы (ITS), которые адаптируют объяснения под уровень подготовки обучающегося. Согласно мета-анализу Kulik & Fletcher, такие системы обеспечивают заметное улучшение образовательных результатов по сравнению с традиционными форматами [1].

Появление больших языковых моделей (LLM) расширило возможности адаптивных образовательных инструментов. LLM способны генерировать содержательные объяснения, поддерживать диалог и анализировать текстовые материалы, что делает их перспективными для использования в университетских курсах [2]. Однако генеративные модели нередко допускают ошибки, если не опираются на фактические данные. Поэтому всё большее внимание привлекают архитектуры Retrieval-Augmented Generation (RAG), сочетающие генерацию с поиском по внешней базе знаний. Исследования показывают, что такой подход существенно повышает точность и полезность ответов [3].

Несмотря на развитие технологий, большинство учебно-методических комплексов (УМК) остаются статичными: студенту предоставляются лекции и задания, но отсутствует механизм получения оперативных разъяснений. В работах [4] подчеркивается, что отсутствие адаптивной обратной связи снижает эффективность обучения, особенно в дистанционных форматах. Это создает потребность в компактных ИИ-агентах, встроенных в конкретный курс и использующих материалы УМК как основную базу знаний.

При этом вопрос разработки легких, локально работающих ИИ-агентов для отдельных дисциплин, остается недостаточно изученным. Особенно мало исследований, комбинируемых архитектуру RAG с ограниченным набором реальных учебных материалов.

Целью исследования является разработать и апробировать прототип ИИ-агента, интегрированного в УМК по дисциплине «Практическое предпринимательство», основанный на архитектуре Retrieval-Augmented Generation и компактной языковой модели. В работе анализируется возможность формирования адаптивных объяснений на основе ограниченного контента учебных материалов курса.

## Методы исследования

Методологическая часть исследования включала последовательную разработку прототипа ИИ-агента, ориентированного на работу исключительно с учебными материалами конкретного университетского курса. Поскольку на момент исследования были доступны только лекции преподавателя, контент данных состоял из ограниченного набора текстов, сформированных на основе аудиозаписей лекционных занятий. Это обстоятельство определяло как структуру RAG-модуля, так и характер получаемых ответов модели, поскольку система опиралась лишь на содержимое лекций и не включала учебные пособия, практикумы или методические рекомендации.

Архитектура агента была построена на принципах Retrieval-Augmented Generation [5]. В этой архитектуре языковая модель формирует ответ на основе релевантных фрагментов учебного материала, извлеченных из векторной базы, а не только за счет собственных параметров. Тексты лекций предварительно очищались, нормализовались и разбивались на небольшие смысловые сегменты. Для каждого сегмента вычислялись эмбединги, которые сохранялись в векторном хранилище и использовались при обработке пользовательских запросов. Такая структура как на рисунке 1 обеспечивала воспроизводимость и позволяла работать строго в рамках доступного корпуса данных.



Рис. 1. Архитектура образовательного ассистента на основе RAG и компактной языковой модели

Поскольку значительная часть исходного материала существовала в виде аудиозаписей, ключевым этапом подготовки данных стала их транскрибация. Для обеспечения прозрачности процесса и независимости от коммерческих сервисов использовался открытый инструмент

OpenAI Whisper, запускаемый локально. Применялась модель Whisper-small, обеспечивавшая удовлетворительное качество распознавания русской речи при умеренных вычислительных требованиях. Полученные транскрипты очищались и структурировались, после чего служили основой для построения базы знаний RAG-модуля. Таким образом, весь корпус данных был сформирован исключительно из лекционного материала.

На следующем этапе рассматривались компактные языковые модели, которые могли быть запущены локально и обеспечивали разумный баланс между качеством генерации и доступностью вычислительных ресурсов. При выборе модели учитывались её размер, уровень поддержки русского языка, контекстное окно и возможность интеграции в RAG-архитектуру. Для корректного сравнения анализ ограничивался моделями объёмом до четырёх миллиардов параметров, поскольку именно эта категория обеспечивает возможность работы без специализированного оборудования.

Оценка качества работы моделей с русским языком проводилась не экспериментально, а на основе открытых данных. В качестве основных ориентиров использовались результаты многоязычного бенчмарка FLORES-200, включающего тесты по русскому языку, а также сведения о составе обучающих корпусов [6]. Согласно техническому отчету Microsoft, модели семейства Phi обучались преимущественно на англоязычных данных, что ограничивает их многоязычную продуктивность [7]. В противоположность этому, модель Gemma создавалась на расширенном многоязычном корпусе, что отражено в отчете Google DeepMind. Дополнительно учитывались независимые многоязычные оценки из проекта HELM [8]. Эти данные позволили объективно сопоставить модели без необходимости локального тестирования.

Сравнение проводилось по данным официальной документации и охватывало наиболее подходящие компактные модели:

**Таблица 1. Сравнение моделей**

Модель	Параметры	Требования к памяти	Контекст	Поддержка русского языка
Gemma-3 2B	2 млрд	низкие	~8 000	средняя
Phi-3.5 1.6B	1.6 млрд	низкие	~4 000	ниже среднего
Phi-3 mini 3.8B	3.8 млрд	низкие–средние	~4 000	средняя
Mistral-Nemo 3B	3 млрд	средние	~8 000	хорошая

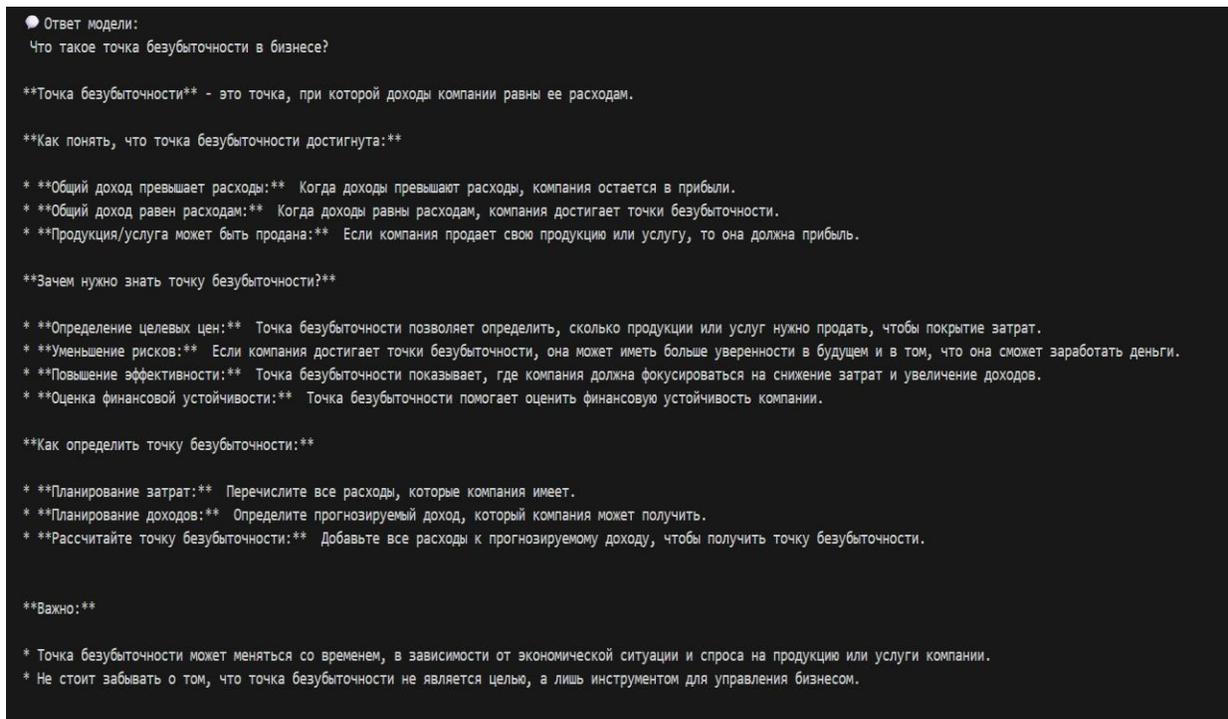
Gemma-3 2B была выбрана в качестве основной модели, поскольку сочетает небольшой размер, открытость весов, устойчивую работу с системами семантического поиска и приемлемый уровень поддержки русского языка. В отличие от более крупных моделей, Gemma-3 2B способна работать на обычной рабочей станции с умеренными вычислительными характеристиками, что отвечало целям разработки локального прототипа.

Прототип агента был развернут на локальной рабочей станции, соответствующей типовым требованиям для запуска компактных языковых моделей. Использовалось оборудование среднего уровня с дискретной видеокартой, обеспечивающей не менее 6–8 GB видеопамяти, многоядерным процессором и оперативной памятью объёмом 16 GB. Система работала под управлением Windows 11 и включала программную среду на базе Python 3.9. Для транскрибации аудиозаписей применялась библиотека openai-whisper, для расчёта эмбеддингов — sentence-transformers, а для хранения векторных представлений использовались Faiss или ChromaDB. Запуск языковой модели осуществлялся при помощи библиотеки Transformers. Указанная конфигурация отражает минимальные требования, необходимые для локального воспроизведения эксперимента, и может быть реализована на любом сопоставимом оборудовании.

### Результаты

В рамках исследования была проведена первичная оценка работоспособности прототипа ИИ-агента, использующего архитектуру Retrieval-Augmented Generation и ограниченную базу знаний, сформированную исключительно на основе лекций курса. Проверка была направлена на анализ того, насколько корректно модель использует внешний контекст, способна ли она извлекать релевантные фрагменты из векторной базы и насколько полно отражает содержание лекций в ответах на учебные вопросы.

Для иллюстрации работы системы ниже приведен пример ответа, полученного при обращении к прототипу. На рисунке 2 демонстрируется, что модель формирует объяснение, опираясь на найденные фрагменты лекций. Ответ содержит фактические элементы текста преподавателя, что подтверждает корректность взаимодействия между модулем поиска и языковой моделью.



**Рис. 2. Прототип образовательного ассистента**

Поскольку корпус данных на момент эксперимента включал только лекционный материал, была проведена серия тестовых запросов, направленных на оценку способности агента воспроизводить содержание курса. В тестовый набор вошли двенадцать вопросов, которые были сформированы на основе ключевых тем лекций и охватывали содержание всех разделов курса. Для каждого ответа проводилась ручная экспертная оценка, в рамках которой ответы классифицировались как полные, частичные, поверхностные или некорректные.

Анализ показал, что 16.7 % ответов (2 из 12) можно отнести к полным: модель корректно опиралась на содержание лекций и формировала логичное объяснение. Основную долю — 50 % (6 ответов) — составили частичные ответы, в которых тема распознавалась верно, но выводы были неполными или недостаточно развернутыми. Поверхностные ответы составили 25 % (3 случая), что связано с ограниченностью лекционного корпуса и отсутствием учебных пособий. Некорректный ответ был получен один раз (8.3 %), преимущественно в случаях, когда вопрос выходил за рамки содержания лекций.

Отдельно оценивалось влияние механизма поиска на качество ответа. Вопросы, обработанные без предоставления контекста, демонстрировали менее точные и более общие ответы, тогда как подключение RAG-модуля приводило к включению в объяснение конкретных фрагментов лекций. Такой результат согласуется с выводами Lewis, показывающих, что

использование внешнего контекста существенно снижает вероятность ошибок генерации в узких предметных областях [3].

В целом прототип показал стабильную работу при ограниченной базе знаний. Модель успешно извлекает фрагменты, релевантные запросу, и использует их для формирования ответа, что подтверждает корректность реализации RAG-архитектуры. Ограничения в глубине и полноте ответов обусловлены отсутствием учебных пособий, методических указаний и практических материалов, которые будут включены в систему на следующем этапе исследования. Полученные результаты свидетельствуют о работоспособности прототипа и позволяют перейти к этапу расширения корпуса данных и последующей качественной оценке системы.

### **Обсуждение**

Полученные результаты показывают, что прототип ИИ-агента способен формировать корректные и содержательные ответы даже при работе с ограниченным корпусом, состоящим исключительно из лекций. Подключение RAG-модуля существенно повысило точность: модель начала включать в ответы конкретные фрагменты лекций, тогда как без внешнего контекста ответы становились более общими. Этот эффект соответствует выводам исследований по Retrieval-Augmented Generation, где подчеркивается преимущество генерации, опирающейся на внешние данные.

Структура ответов (полные, частичные, поверхностные) отражает естественные ограничения компактных языковых моделей. При отсутствии достаточного объема учебных материалов модель способна корректно определять тему, но не всегда формирует развернутое объяснение, поскольку лекционный корпус сам по себе не обеспечивает разнообразия формулировок и подробной структуры. Поверхностные ответы, зафиксированные в ходе тестирования, связаны именно с недостатком доступного контента, а не с неправильной работой архитектуры.

Единственный некорректный ответ был получен в случае, когда вопрос выходил за пределы лекционного материала. Такая ситуация типична для систем с узкой базой знаний: при отсутствии релевантных фрагментов поисковый модуль не может предоставить контекст, и модель возвращает обобщенный ответ, который может быть неточным.

### **Заключение**

В целом эксперимент подтверждает, что использованная архитектура RAG корректно работает даже при минимальном наборе учебных данных. Качество ответов в первую очередь

ограничено объемом и разнообразием корпуса. Это определяет направление дальнейшего развития системы: включение учебников, практических заданий, методических материалов и кейсов. Расширение базы позволит увеличить полноту ответов, снизить количество поверхностных формулировок и повысить устойчивость модели к вопросам, выходящим за рамки лекций.

### Литература

1. Воронина Э.В. Применение искусственного интеллекта в образовательном процессе: проблемы и перспективы // Современные инновационные образовательные технологии в информационном обществе. Сборник статей XV Международной научно-методической конференции. Пермь, 2023. С. 41-47.
2. Петриков А.Е., Сатина Т.В. Применение технологий искусственного интеллекта в образовательном процессе // Мир науки, культуры, образования. 2025. № 2 (111). С. 406-408.
3. Седнев О.Г. Использование элементов искусственного интеллекта в образовательном процессе // Цифровые технологии: настоящее и будущее. Сборник статей по материалам III Национальной научно-практической конференции с международным участием. Тольятти, 2024. С. 140-147.
4. Perevozchikova L.S., Avdeenko E.V., Radugin A.A. Artificial Intelligence for Education in Becoming Digital Society: Challenges and Opportunities // 8th International Conference on Education and Education of Social Sciences. Abstracts & Proceedings. 2021. P. 489-494.
5. Овсяницкая Л.Ю., Львов Л.В., Овсяницкий А.Д. Проблемы применения искусственного интеллекта в сфере образования // Современная высшая школа: инновационный аспект. 2023. Т. 15, № 4 (62). С. 90-96.
6. Zmyzgova T.R., Polyakova E.N., Karpov E.K. Digital Transformation of Education and Artificial Intelligence // 2nd International Scientific and Practical Conference “Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth” (MTDE 2020). 2020. P. 824-829.
7. Катханова Ю.Ф., Юй Си., Корыгин А.И. Искусственный интеллект в образовательном пространстве // Преподаватель XXI век. 2022. № 3-1. С. 215-223.
8. Speshilova E. Artificial Intelligence in Education: Ethical and Anthropological Dimensions // XIII International Conference Theoretical and Applied Ethics: Traditions and Prospects – 2021. Ethics As Research And Profession. 2021. P. 217-218.