

УДК 338.31

### 33. СИСТЕМА ОПТИМИЗАЦИИ РАБОТЫ HR-ОТДЕЛА

*Худницкий А.В., студент гр.353502, Шауро Е.А., студент гр.353502  
Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Насонова И.В. – кандидат экономических наук*

**Аннотация.** В работе рассматривается концепт внедрения корпоративной RAG-системы (Retrieval-Augmented Generation) на базе локальных нейросетевых моделей для автоматизации работы HR-отделов. Анализируется проблема высокой нагрузки на HR-специалистов рутинными запросами сотрудников. Проводится SWOT-анализ и оценка экономической эффективности проекта. Предложена архитектура On-Premise решения с интеграцией базы знаний, встроенными метриками качества и дашбордами бизнес-аналитики. Акцент сделан на информационную безопасность и окупаемость технологии.

**Ключевые слова.** RAG-система, локальные LLM, искусственный интеллект, HR-менеджмент, корпоративные базы знаний, Notion, автоматизация бизнес-процессов, чат-бот, векторные базы данных.

Искусственный интеллект как научное направление существует уже достаточно давно: концепции машинного обучения и первые нейросети начали зарождаться еще во второй половине прошлого века, а первый виртуальный собеседник (чат-бот ELIZA) был разработан в 1966 году [1]. В конце 2022 – начале 2023 года сфера обрела большую популярность и перешла в фазу ускоренного роста с появлением третьей модели ChatGPT компании OpenAI [2]. Секрет этого прорыва заключался не в создании принципиально новых алгоритмов, а в успешной упаковке сложной языковой модели в интуитивно понятный формат диалогового окна, что устранило технические барьеры и демократизировало доступ к ИИ, позволив взаимодействовать с ним миллионам обычных пользователей.

Сегодня искусственный интеллект используется практически везде, став неотъемлемой частью технологического стека большинства компаний. По данным глобальных исследований, в 2023 году 92% компаний планировали увеличить свои инвестиции в искусственный интеллект в ближайшие три года [3]. Форматы использования ИИ вышли далеко за пределы базовых чат-ботов: сегодня современный рынок активно применяет генераторы программного кода, инструменты для создания изображений, предиктивную аналитику, а также специализированные решения для медицины, логистики и HR.

На фоне повсеместного внедрения больших языковых моделей (LLM) особую нишу заняли системы дополненной генерации (Retrieval-Augmented Generation, или RAG), которые решают ключевые ограничения базовых нейросетей.

RAG-архитектура объединяет два основных компонента, а именно механизм информационного поиска (как правило, векторные базы данных, хранящие информацию в виде эмбедингов) и саму генеративную языковую модель. При поступлении запроса система сначала извлекает наиболее релевантные фрагменты текста из внешней базы знаний, а затем передает их LLM для формирования ответа. Кроме того, RAG решает проблему зависимости от статичных обучающих данных, так как жестко привязывает ответы нейросети к проверенным внешним источникам и позволяет предоставлять пользователю ссылки на конкретные документы для верификации. Благодаря RAG компании способны быстро разворачивать корпоративные базы знаний и службы поддержки, способные генерировать безошибочные ответы со специфичной для конкретного бизнеса информацией.

Именно в свете высокой актуальности и востребованности подобных решений, главной целью настоящей статьи и представляемого в рамках конференции проекта является разработка концепта локальной корпоративной RAG-системы. Данное решение направлено на комплексную оптимизацию работы HR-отделов, так как оно призвано автоматизировать рутинные процессы консультирования сотрудников, вывести взаимодействие с внутренней документацией на качественно новый уровень и при этом гарантировать абсолютную информационную безопасность данных компании.

Смотря на конкурентов и особенности внедрения RAG-систем, видно, как корпоративный бизнес меняет подходы к созданию ИИ-инструментов. Главный стратегический вопрос, стоящий перед бизнесом сегодня – это дилемма «Build vs. Buy», что означает создавать систему внутри силами собственного IT-отдела или покупать готовый внешний продукт с рынка.

Согласно масштабному исследованию венчурного фонда Menlo Ventures [4], соотношение «Build vs. Buy» на корпоративном рынке ИИ полностью перевернулось за период 2024 - 2025 годов: если в 2024 году 47% решений разрабатывались внутри компаний, а оставшиеся 53% закупались, то в следующем году лишь 24% RAG-систем создавались силами внутренних IT-отделов, в то время как 76% внедрений представляли собой покупку готовых платформ и SaaS-приложений.

Приведенная статистика обусловлена серьезными барьерами, с которыми сталкивается бизнес при попытке самостоятельной разработки RAG-архитектуры. Ниже отмечены основные из них.

1 Финансовые и временные затраты. По оценкам Gartner и Dimensional Research [5], создание качественной in-house RAG-системы «с нуля» обходится предприятиям в сумму от 750 000 до 1 000 000

долларов США. При этом у 88% компаний уходит более 6 месяцев только на запуск первого рабочего прототипа.

2 Дефицит кадров. Около 51% бизнеса испытывают острую нехватку AI-специалистов [6], а поддержка локальной распределенной системы требует выделенной команды из Data-инженеров и ML-разработчиков.

3 Проблема масштабирования. Исследования показывают, что 61% компаний, разрабатывающих RAG самостоятельно, сталкиваются с критическими проблемами точности («галлюцинациями»), и менее 10% проектов успешно выходят из стадии Proof of Concept в полноценное использование.

Поэтому, внешние B2B-продукты захватили рынок благодаря высокой скорости развертывания (дни вместо месяцев) и наличию готовых коннекторов к популярным CRM и ERP системам.

Однако для HR-отделов использование коммерческих облачных SaaS-решений создает критический риск утечки персональных и корпоративных данных, например NDA, штатных расписаний, заработных плат, а стоимость enterprise-лицензий остается высокой. В качестве альтернативы, объединяющей безопасность разработки и скорость внедрения готовых продуктов, в рамках данного исследования был разработан рабочий прототип полностью локальной корпоративной RAG-системы [7].

Техническая архитектура прототипа базируется на open-source компонентах и библиотеке оркестрации LangChain. В качестве базы знаний используется рабочее пространство Notion, откуда HR-документация и регламенты бесшовно выгружаются через API. Для высокоточного извлечения информации реализован гибридный поиск, объединяющий векторную базу данных FAISS и лексический алгоритм BM25. За генерацию итоговых ответов несет ответственность локально развернутая большая языковая модель Ollama. Это означает, что вся обработка текстов происходит строго на серверах компании, гарантируя 100% конфиденциальность данных.

Особое внимание в проекте уделено пользовательскому опыту. Взаимодействие сотрудников с ИИ-ассистентом реализовано через популярный мессенджер Telegram [8]. Такая интеграция обеспечивает максимальное удобство, потому как сотрудникам не требуется осваивать новый софт или устанавливать сторонние приложения, а сам чат-бот предоставляет интуитивно понятный интерфейс с поддержкой сохранения контекста диалога и кнопками для быстрой оценки качества ответа, что проиллюстрировано на рисунке 1. Дополнительно в систему встроен программный модуль метрик, который программно проверяет каждый ответ на достоверность и полноту.

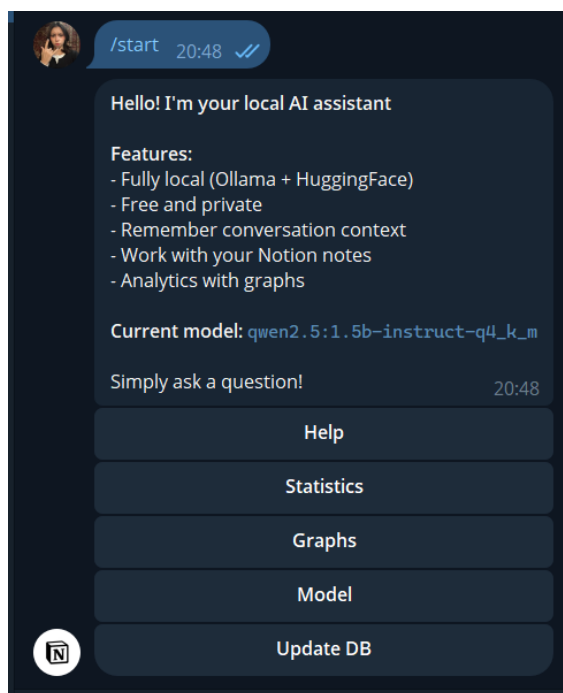


Рисунок 1 – Интерфейс пользователя

Данный технологический стек формирует систему, у которой можно выделить следующие конкурентные преимущества: высокую безопасность, удобство масштабирования системы и отсутствие регулярных платежей за использование API сторонних нейросетей.

Однако с этими сильными и положительными сторонами сопряжены также и риски, рассмотренные в матрице SWOT-анализа на рисунке 2 [9].



Рисунок 2 – Матрица SWOT-анализа предлагаемого проекта

Анализ сильных сторон и возможностей отражает высокий потенциал предлагаемого решения. RAG-подход дает устойчивое преимущество перед конкурентами, поскольку система обеспечивает безопасность данных и снижает затраты на вычисления. Растущий интерес бизнеса к ИИ-решениям и развитие открытых языковых моделей создают условия для своевременного выхода продукта на рынок и его последующего масштабирования за пределы HR-функции.

Вместе с тем выявленные слабые стороны носят преимущественно операционный характер и поддаются управлению. Зависимость качества системы от полноты корпоративной базы знаний и требования к IT-инфраструктуре могут быть нивелированы за счет поэтапного внедрения, а именно первоначального аудита и структурирования документации, с последующей внешней технической поддержкой на старте проекта.

Наиболее значимой из выявленных угроз является риск генерации некорректных ответов в особо значимых HR-контекстах, однако он может существенно снижаться самой природой RAG-архитектуры [10]. Она жестко привязывает ответы модели к верифицированным источникам, и, что главное, предоставляет ссылки на конкретные документы. Угроза сопротивления персонала преодолевается изменением позиционирования системы: она представляется не как замена HR-специалиста, а как инструмент освобождения его от рутинных запросов в пользу задач более высокой ценности.

В совокупности результаты SWOT-анализа свидетельствуют о том, что риски проекта управляемы, а соотношение сильных сторон и возможностей создает прочную основу для успешной реализации спроектированного решения.

Разработка экономического обоснования для корпоративной RAG-системы требует комплексного анализа затрат на внедрение и последующей оценки достигаемого эффекта. Представленный ниже анализ базируется на среднерыночных значениях для белорусского рынка и ориентирован на компанию среднего размера (150-500 сотрудников, 5-8 HR-специалистов) [11]. Таблица 1 детализирует структуру затрат, необходимых для запуска и функционирования системы, с разделением на капитальные (приобретение сервера, разработка) и текущие (поддержка, обновление базы знаний) расходы.

Таблица 1 – Статьи расходов для предлагаемой RAG-системы в год, тыс. бел. руб.

| Статья расходов     | Приобретение сервера | Разработка и настройка системы | Наполнение базы знаний | Поддержка и обновление системы | Обновление базы знаний | Итого |
|---------------------|----------------------|--------------------------------|------------------------|--------------------------------|------------------------|-------|
| При запуске проекта | 10,5                 | 5                              | 2                      | –                              | –                      | 17,5  |
| В год               | –                    | –                              | –                      | 2,4                            | 0,6                    | 3     |

Прогнозируемые экономические показатели, представленные в таблице 2, рассчитаны исходя из следующих допущений: средняя заработная плата HR-специалиста – 2 305 BYN/мес. [12], доля рабочего времени, занятая рутинными запросами сотрудников – 35% [13][14], ожидаемое высвобождение этого времени после внедрения системы – 80% (система берет на себя типовые обращения). При отделе из 6 HR-специалистов годовая экономия на фонде оплаты труда составляет порядка 59 348 BYN за счет перераспределения высвобожденного времени на задачи более высокой ценности либо сокращения потребности в дополнительном найме.

Таблица 2 – Экономические показатели для предлагаемой RAG-системы

| Экономический показатель | Годовая экономия          | Совокупные затраты за 3 года | Чистая прибыль за 3 года | ROI за 3 года | Срок окупаемости |
|--------------------------|---------------------------|------------------------------|--------------------------|---------------|------------------|
| Значение                 | 46565 тыс. бел. руб в год | 26500 тыс. бел. руб          | 113195 тыс. бел. руб     | 427%          | 4.5 месяца       |

Положительное значение чистой прибыли за трехлетний период подтверждает экономическую целесообразность проекта. Высокий показатель ROI (427%) объясняется относительно низкой стоимостью внедрения на базе open source компонентов по сравнению с коммерческими аналогами, стоимость enterprise-лицензий которых достигает десятков тысяч долларов в год. Срок окупаемости около 4.5 месяцев является весьма благоприятным для проектов подобного класса [15].

В условиях, когда специалисты по управлению персоналом тратят до 35% рабочего времени на обработку типовых запросов, а использование облачных ИИ-сервисов сопряжено с критическими рисками утечки конфиденциальных данных, предложенное решение демонстрирует оптимальный баланс между инновационностью и строгой информационной безопасностью.

Внедрение RAG является значительно более рентабельным и быстрым подходом по сравнению с ресурсоемким процессом постоянного дообучения моделей под новые корпоративные данные. Система передает нейросети только самые релевантные фрагменты (чанки) справочного материала, что снижает размер запросов, повышает скорость обработки и экономит вычислительные мощности. Кроме того, наличие встроенного модуля метрик качества позволяет контролировать достоверность генерируемых ответов и минимизировать риск ИИ-«галлюцинаций», что абсолютно критично при работе с значимой кадровой документацией.

Экономическое обоснование подтверждает коммерческую привлекательность разработанного продукта. Расчеты показывают, что для компании среднего размера стартовые инвестиции в локальную инфраструктуру и настройку системы окупаются всего за 4.5 месяца, обеспечивая впечатляющий показатель ROI в 427% за трехлетний период. Делегирование ИИ-ассистенту до 80% рутинных обращений генерирует существенную экономию фонда оплаты труда. Однако главная ценность проекта заключается не только в прямом снижении издержек, но и в трансформации роли HR-отдела: освобожденные человеческие ресурсы могут быть перенаправлены на решение стратегических задач – развитие талантов, онбординг, укрепление корпоративной культуры и HR-бренда компании.

#### Список использованных источников:

1. Weizenbaum, J. ELIZA – a computer program for the study of natural language communication between man and machine / J. Weizenbaum // Communications of the ACM. – 1996. – Vol. 9, № 1. – P. 36 - 45.
2. Language Models are Few-Shot Learners [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2005.14165>.
3. The state of AI in 2023: Generative AI's breakout year [Электронный ресурс]. – Режим доступа: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year>.
4. The great AI flip: Why 76% of enterprises stopped building AI in-house [Электронный ресурс]. – Режим доступа: <https://beam.ai/agentics-insights/the-great-ai-flip-why-76-of-enterprises-stopped-building-ai-in-house>.
5. Beyond 'build vs. buy': Why a unified AI platform is the only way to scale agentic AI [Электронный ресурс]. – Режим доступа: <https://writer.com/blog/build-vs-buy-generative-ai/>.
6. Build vs Buy AI: Decision Framework for Smart Implementation [Электронный ресурс]. – Режим доступа: <https://writer.com/blog/build-vs-buy-generative-ai/>.
7. HR department optimization system [Электронный ресурс]. – Режим доступа: [https://github.com/AlHudnitskii/rag\\_notion](https://github.com/AlHudnitskii/rag_notion).

8. Telegram Bot Docs [Электронный ресурс]. – Режим доступа: <https://core.telegram.org/bots/api>.
9. What is SWOT-analysis? [Электронный ресурс]. – Режим доступа: <https://core.telegram.org/bots/api>.
10. RAG Architecture [Электронный ресурс]. – Режим доступа: <https://incom.by/articles/resheniya-dlya-sovmestnoy-raboty/arkhitektura-rag-tekhnicheskiy-obzor/>.
11. Optimal amount of HR-department [Электронный ресурс]. – Режим доступа: <https://belarus.gorodrabot.by/hr-менеджер>.
12. Vacancies of HR-manager in Belarus [Электронный ресурс]. – Режим доступа: [https://www.profiz.ru/stup/7\\_2021/skolko\\_hr/](https://www.profiz.ru/stup/7_2021/skolko_hr/).
13. Stratus HR. Estimated Time On HR: Employee Relations & Compliance Tasks [Электронный ресурс]. – Режим доступа: <https://stratus.hr/resources/estimated-time-on-hr-employee-relations-compliance-tasks>.
14. Deskpro. HR Helpdesk Analytics: The 5 Critical Metrics That Drive Employee Satisfaction & Prove HR Value [Электронный ресурс]. – Режим доступа: <https://www.deskpro.com/blog/hr-help-desk-metrics>.
15. RAG in corporate products [Электронный ресурс]. – Режим доступа: <https://www.it-world.ru/cionews/nefa7mi259cggw00swwsos8scsg0k08.html>.

UDC 338.31

## HR-DEPARTMENT OPTIMIZATION SYSTEM

*Khudnitskii A.V., student of group 353502, Shauro E.A., student of group 353502*

*Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

*Nasonova I.V. – candidate of Economic Sciences*

**Annotation.** The paper considers the concept of implementing a corporate RAG system (Retrieval-Augmented Generation) based on local neural network models for automating the work of HR departments. The problem of high workload on HR specialists by routine requests of employees is analyzed. A SWOT analysis and an assessment of the economic efficiency of the project is carried out. The architecture of an On-Premise solution with knowledge base integration, built-in quality metrics, and business intelligence dashboards is proposed. The emphasis is on information security and technology payback.

**Keywords.** RAG-system, local LLM, artificial intelligence, HR-management, corporate knowledge bases, Notion, business-process automation, chat-bot, vector databases.