

ЛОКАЛЬНЫЙ СЕРВЕР ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ЦИФРОВЫЕ АССИСТЕНТЫ

А. В. МАТВЕЕВ

Ключевые слова: интеллектуальная поддержка принятия решений; база знаний; база данных; интеллектуальный поиск; веб-ИИ-агенты; большие языковые модели; локальный сервер искусственного интеллекта; управление.

Keywords: intelligent decision support; knowledge base; database; intelligent search; web AI agents; large language models; local AI server; management.

Аннотация. Актуальность исследования и программной реализации обусловлена необходимостью повышения оперативности, обоснованности и эффективности управленческих решений в современных условиях. Существующие подходы, основанные преимущественно на экспертных методах, не всегда позволяют своевременно и точно учитывать весь спектр факторов, влияющих на выбор управленческих решений, что требует разработки и внедрения структурных компонентов интеллектуальных систем поддержки принятия решений.

Abstract. The relevance of the study and program implementation is due to the need to improve the efficiency, validity and effectiveness of management decisions in modern conditions. Existing approaches, based primarily on expert methods, do not always allow for a timely and accurate consideration of the entire range of factors influencing the choice of management decisions, which requires the development and implementation of elements of intelligent decision support systems.

Введение. Цифровизация и искусственный интеллект становятся важнейшими элементами долгосрочной политики России, направленной на устойчивый экономический рост, модернизацию ключевых отраслей, повышение эффективности управления и технологическую независимость [1, 2]. С каждым днем растет востребованность в углубленных научных знаниях по многогранным аспектам развития ИИ. Стремительно развиваются наиболее совершенные технологии искусственного интеллекта, такие как большие языковые модели, интеллектуальные системы управления и поддержки принятия решений, генеративный ИИ [3, 4]. Первоосновой является создание собственных конкурентоспособных моделей ИИ, а также специализированных агентов – автономных интеллектуальных программ, которые могут взаимодействовать с внешней средой, принимать решения и предпринимать определенные действия без вмешательства человека.

Материалы и методы. Работы проведены на основе применения цифровых технологий, включая методы, алгоритмы и большие языковые модели искусственного интеллекта в сочетании с инновационным подходом. При разработке использовано свободно распространяемое кроссплатформенное программное обеспечение с открытым исходным кодом.

Результаты и обсуждение. Запуск моделей и агентов ИИ на собственных вычислительных мощностях, без использования облачных платформ, обладает множеством преимуществ: контроль над данными и их безопасность (исключается сценарий, когда модель станет источником утечки частных данных или плагиата защищенного авторским правом контента), снижение затрат на облачные вычисления и возможность использования высокопроизводительных локальных ресурсов для ускорения

обучения и получения выводов/прогнозов. Локальный запуск обеспечивает гибкость и возможность тонкой настройки моделей ИИ, позволяя исследователям настраивать архитектуры и экспериментировать с гиперпараметрами, чтобы повысить точность и глубину автоматической генерации контента. Кроме того, локальная среда удобна для разработки и тестирования, позволяет быстро вносить изменения и проводить эксперименты, а также исключает зависимость от интернет-соединения, что особенно полезно в условиях ограниченного доступа к сети, в случае блокировки, санкций, удаления модели разработчиками из общего доступа.

В 2025 г. в ФГБНУ «ФНЦ ВНИИГиМ им. А.Н. Костякова» выполнено развертывание локального сервера моделей искусственного интеллекта. Ключевые программные составляющие данного сервера:

1. Ollama – фреймворк для локального запуска и управления большими языковыми моделями;
2. Windows Subsystem for Linux – подсистема совместимости для запуска Linux-приложений в операционных системах Windows 10 и 11;
3. Docker – платформа, позволяющая инкапсулировать приложения и все зависимости в изолированные контейнеры;
4. Open WebUI – веб-интерфейс, предназначенный для работы с различными LLM-интерфейсами, такими как Ollama. Предлагает широкий спектр функций – основные из них направлены на повышение удобства в управлении моделями и запросах к ним.

Ollama, Docker и Open WebUI являются свободно распространяемым кроссплатформенным программным обеспечением с открытым исходным кодом. Большие языковые модели загрузились в Ollama из общедоступного онлайн-хранилища Huggingface (<https://huggingface.co>).

На высокоуровневом языке Python автором разработан веб-ассистент для организации интеллектуального поиска по базе знаний с помощью применения больших языковых моделей, что позволяет обеспечить полноту поиска и релевантность [5]. Программа реализована в виде диалогового агента, обладающего веб-интерфейсом, с которым можно общаться в формате «вопрос-ответ», и системы генерации с дополненной выборкой (Retrieval Augmented Generation).

Генерация с дополненной выборкой – сложная система, разработанная для расширения возможностей больших языковых моделей за счет интеграции мощных механизмов поиска информации. Представляет собой сочетание традиционных языковых моделей с инновационным подходом: в процесс генерации напрямую интегрируется механизм поиска информации. ИИ перед тем, как ответить на вопрос, может обратиться к библиотеке текстов, что делает его более осведомленным и контекстно точным (рис. 1).

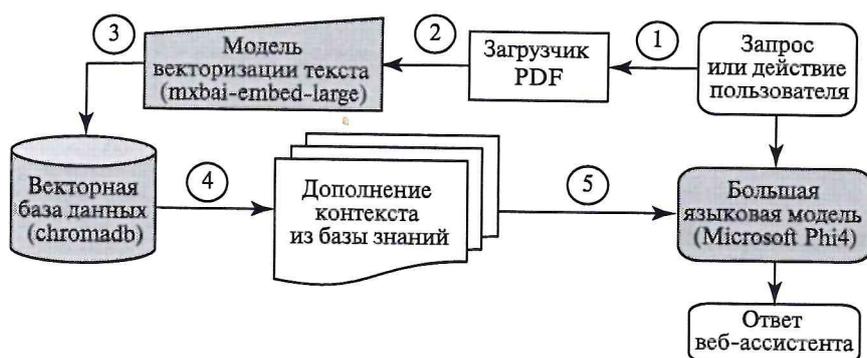


Рис. 1. Алгоритм работы разработанного веб-ассистента для интеллектуального поиска по базе знаний

Программные компоненты, которые легли в основу разработанного веб-ассистента для интеллектуального поиска по базе знаний:

1. Ollama – локальный сервер моделей ИИ;
2. Langchain – платформа, которая помогает облегчить интеграцию больших языковых моделей;
3. Phi4 – большая языковая модель от Microsoft;
4. mxbai-embed-large – мультиязычная модель для векторизации текста (представление текста в виде массива чисел) от проекта mixedbread.ai;
5. PyPDFLoader – загрузчик и обработчик PDF-файлов;
6. ChromaDB – векторная база данных для разработки приложений ИИ;
7. Набор Python-библиотек для создания веб-приложений.

На вход веб-ИИ-ассистента пользователем подается база знаний в формате PDF. Это может быть монография, диссертация, отчет НИР, подборка научных статей и т. д. Программа обрабатывает загруженный PDF-файл, с помощью многоязычной модели векторизации преобразовывает текст в набор числовых векторов, полученное векторное представление заносит в векторную базу данных, дополненный контекст из базы знаний передает на вход большой языковой модели, которая отвечает на поставленный пользователем вопрос. Поддерживается многопользовательская работа, каждый пользователь может загружать свою базу знаний и формировать свои вопросы.

На рис. 2 приведен фрагмент диалога с веб-ИИ-ассистентом после загрузки монографии [6] (здесь и далее стоит отметить, при разработке веб-ИИ-ассистентов автором использовались большие языковые модели общего назначения. По мере появления в свободном доступе наукоемких, специализированных моделей качество генерируемых ответов будет расти):

Мой вопрос: Какие наблюдения проводились в вегетационный период?

Ответ ИИ: Из предоставленного контекста следует, что в вегетационный период был заложен опыт по применению удобрительного биокомпоста для восстановления плодородия деградированной аллювиальной почвы.

В главе 3.3 «Результаты вегетационного эксперимента по применению удобрительного биокомпоста для восстановления плодородия деградированной аллювиальной почвы» описываются результаты данного опыта.

Вероятно, в вегетационный период проводились следующие наблюдения:

Изучение влияния биокомпоста на урожайность и качество сельскохозяйственных культур;

Изменение показателей плодородия почвы (содержание гумуса, питательных веществ и т. д.);

Динамика роста и развития растений;

Биологическая активность и микробиологические процессы в почве.

Точные детали методики наблюдений и показателей, которые изучались, возможно, описаны в главе 3.2 «Условия и методика проведения экспериментальных исследований».

Автором также разработан интеллектуальный веб-агент для анализа баз данных (либо структурированных наборов данных) с помощью применения больших языковых моделей [7]. Программа также реализована в виде диалогового ассистента, обладающего веб-интерфейсом, с которым можно общаться на естественном языке (рис. 3). На вход программы пользователем подается база данных в формате CSV (текстовый формат, предназначенный для представ-

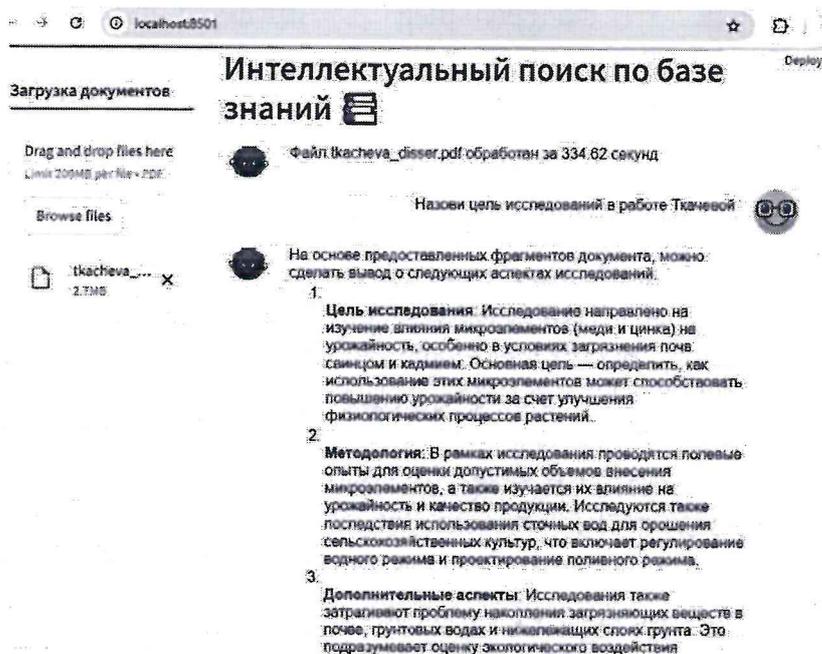


Рис. 2. Веб-ИИ-ассистент отвечает на вопросы по загруженной кандидатской диссертации

ления табличных данных; поддерживается Excel, Access, MySQL и т. д.) и задается интересующий вопрос. Например: «Предоставь сводку о концентрации тяжелых металлов в донных отложениях реки Донховка: cd: [@cd], pb: [@pb], hg: [@hg], zn: [@zn], cr: [@cr], co: [@co], ni: [@ni], cu: [@cu]».

Программа обрабатывает загруженный CSV-файл с результатами отбора проб донных отложений водных объектов, подключается к локальному серверу искусственного интеллекта, посредством запросов POST построчно передает данные на вход выбранной пользователем большой языковой модели (например, Microsoft Phi4, Google Gemma3, Deepseek и т. д.), которая выполняет суммаризацию текста и резюмирование данных в режиме реального времени. Фрагмент вывода работы ИИ-агента:

16, Река Донховка, 0.28, 163, 0.031, 1055, 262, 10.1, 11, 1, 43.3, «Вот сводка о концентрации тяжелых металлов в донных отложениях реки Донховка:

1. ****Цинк (Zn)**:** 1055 мг/кг

– Цинк является одним из наиболее обильных элементов в донных отложениях. Его высокая концентрация может быть связана с естественными источниками, такими как выветривание пород, а также антропогенными источниками, такими как промышленные выбросы и сельскохозяйственные удобрения.

2. ****Хром (Cr)**:** 262 мг/кг

– Хром часто встречается в окружающей среде из-за промышленных процессов, таких как производство стали и красок. Его присутствие в донных отложениях может указывать на промышленные выбросы или использование хромосодержащих материалов.

3. ****Свинец (Pb)**:** 163 мг/кг

– Свинец является значительной проблемой из-за его токсичности. Его высокая концентрация может быть связана с историческим использованием свинцовой краски, бензина и промышленных выбросов.

Закключение. Проведенные исследования позволили реализовать возможность тонкой настройки гиперпараметров большой языковой модели. Так можно задавать: степень случайности в ответах модели; методы выборки токенов при генерации текста; штрафы за частоту и присутствие слов в ответе для расширения спек-

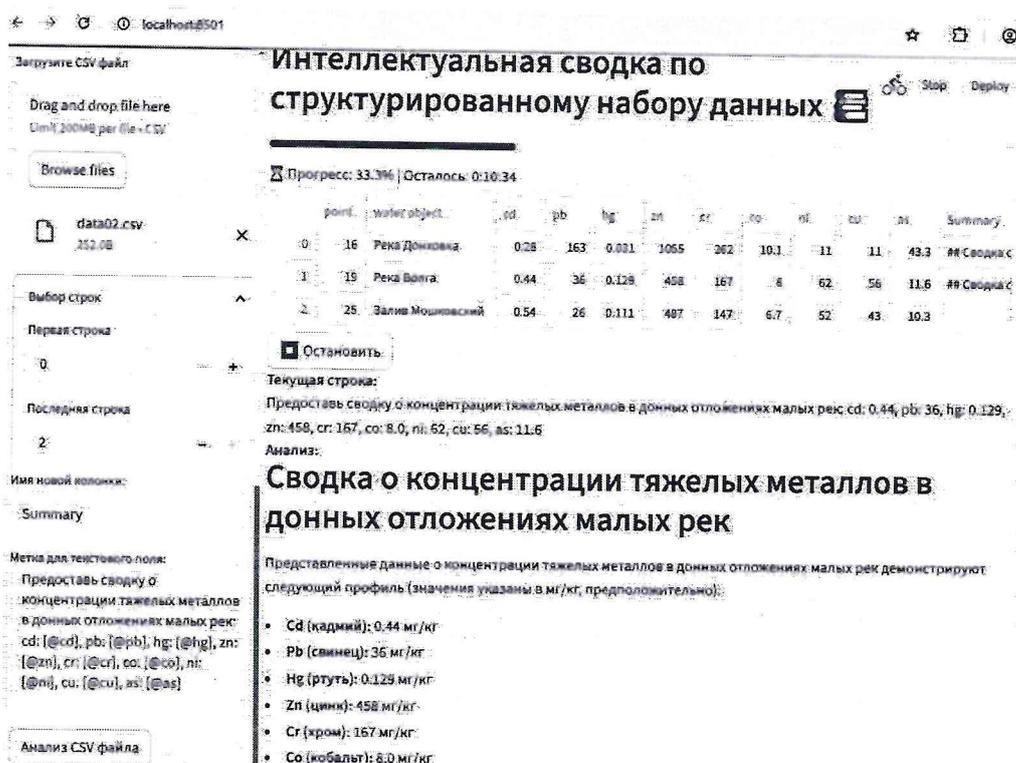


Рис. 3. Веб-ИИ-агент для анализа структурированного набора данных

тра рассматриваемых моделью тем и увеличения разнообразия в выборе слов; максимальное количество токенов, которое разрешено сгенерировать; размер контекстного окна модели и т. д. Кроме того, можно установить свой системный промпт – руководящие инструкции для модели, правильный выбор которых позволит значительно повысить качество и релевантность генерируемых текстов. Результат работы веб-ИИ-агента можно скопировать в буфер обмена, либо сохранить в виде структурированного аналитического отчета.

В настоящее время существует множество больших языковых моделей, которые можно развернуть на локальном компьютере и пользоваться всеми возможностями искусственного интеллекта без доступа в Интернет. При этом можно оптимизировать модель, чтобы повысить точность генерации ответов, и настроить агентов для выполнения определенных операций под конкретные задачи. Представленные веб-ИИ-агенты являются современным адаптивным инструментом, позволяющим создавать базы знаний с интеллектуальным поиском и анализировать базы данных.

На агропромышленной выставке «Золотая осень 2025» разработана «Интеллектуальная мультиагентная геоинформационно-аналитическая система мониторинга, прогнозирования и поддержки принятия решений для совершенствования организации управления мелиоративно-водохозяйственным комплексом» (авторский коллектив: академик РАН В.А. Шевченко, к.т.н. А.В. Матвеев, д.т.н. С.Д. Исаева, д.с.-х.н. Э.Б. Дедова, к.с.-х.н. А.А. Дедов), включающая описанные веб-ИИ-агенты, удостоена серебряной медали.

ЛИТЕРАТУРА

1. О развитии искусственного интеллекта в Российской Федерации (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»): Указ Президента РФ от 10.10.2019 № 490 (ред. от 15.02.2024) URL: https://www.consultant.ru/document/cons_doc_LAW_335184/ (дата обращения 23.07.2025)/
2. Заседание Совета по стратегическому развитию и национальным проектам. Стенограмма 21.12.2023 [Электронный ресурс]. URL: <http://www.kremlin.ru/events/councils/by-council/1029/73083> (дата обращения 23.07.2025)/
3. Интеллектуальная система поддержки принятия решений для управления мелиоративно-водохозяйственным комплексом. К 100-летию Всероссийского научно-исследовательского института гидротехники и мелиорации им. А.Н. Костякова / С.Д. Исаева, Э.Б. Дедова, А.В. Матвеев [и др.]. М.: ФНЦ ВНИИГиМ им. А.Н. Костякова, 2024. 204 с. ISBN 978-5-907464-78-0. DOI 10.37738/VNIIGIM.2024.19.62.001. EDN ZTXNHM.
4. Веб-ГИС-платформа для обоснования решений по управлению мелиоративно-водохозяйственным комплексом Республики Калмыкия: Свидетельство о государственной регистрации программы для ЭВМ № 2024687737 Российская Федерация. № 2024686384: заявл. 06.11.2024; опублик. 20.11.2024 / А.В. Матвеев, Э.Б. Дедова, С.Д. Исаева, Р.М. Шабанов; заявитель ФНЦ ВНИИГиМ им. А.Н. Костякова. EDN GWNAHL.
5. Веб-ассистент для интеллектуального поиска по базе знаний: Свидетельство о государственной регистрации программы для ЭВМ № 2025618901 Российская Федерация: заявл. 28.03.2025; опублик. 08.04.2025 / А.В. Матвеев; заявитель ФНЦ ВНИИГиМ им. А.Н. Костякова. EDN ESBIWC.
6. Новые способы повышения продуктивности деградированных мелиорированных земель с применением информационных технологий / А.В. Ильинский, Н.В. Коломийцев, А.В. Матвеев [и др.]. М.: ВНИИГиМ им. А.Н. Костякова, 2022. 152 с. ISBN 978-5-907464-34-6. EDN BVVIEO.
7. Интеллектуальный веб-агент для анализа баз данных: Свидетельство о государственной регистрации программы для ЭВМ № 2025682023 Российская Федерация: заявл. 29.07.2025; опублик. 20.08.2025 / А.В. Матвеев; заявитель ФНЦ ВНИИГиМ им. А.Н. Костякова. EDN FUDKJJ.

REFERENCES

1. Decree of the President of the Russian Federation of 10.10.2019 No. 490 (as amended on 15.02.2024) «On the Development of Artificial Intelligence in the Russian Federation» (together with the

«National Strategy for the Development of Artificial Intelligence for the Period up to 2030»). URL: https://www.consultant.ru/document/cons_doc_LAW_335184/ (date of access 23.07.2025).

2. Meeting of the Council for Strategic Development and National Projects. Transcript of 21.12.2023 [Electronic resource]. URL: <http://www.kremlin.ru/events/councils/by-council/1029/73083> (date of access 23.07.2025).

3. Intelligent decision support system for managing the land reclamation and water management complex: On the 100th anniversary of the All-Russian Research Institute of Hydraulic Engineering and Land Reclamation named after A.N. Kostyakov / S.D. Isaeva, E.B. Dedova, A.V. Matveev [et al.]. Moscow: All-Russian Research Institute of Hydraulic Engineering and Land Reclamation named after A.N. Kostyakov, 2024. 204 p. ISBN 978-5-907464-78-0. DOI 10.37738/VNIIGIM.2024.19.62.001. EDN ZTXNHM.

4. Certificate of state registration of computer program No. 2024687737 Russian Federation. Web GIS platform for substantiating decisions on managing the melioration and water management complex of the Republic of Kalmykia: No. 2024686384: declared. 06.11.2024; published. 20.11.2024 / A.V. Matveev, E.B. Dedova, S.D. Isaeva, R.M. Shabanov; applicant Federal State Budgetary Scientific Institution «Federal Scientific Center for Hydraulic Engineering and Land Reclamation named after A.N. Kostyakov». EDN GWNAHL.

5. Certificate of state registration of computer program No. 2025618901 Russian Federation. Web assistant for intelligent search in the knowledge base: declared 28.03.2025; published 08.04.2025 / A.V. Matveev; applicant Federal State Budgetary Scientific Institution «Federal Scientific Center for Hydraulic Engineering and Land Reclamation named after A.N. Kostyakov». EDN ESBIWC.

6. New methods for increasing the productivity of degraded reclaimed lands using information technologies / A.V. Ilyinsky, N.V. Kolomyitsev, A.V. Matveev [et al.]. Moscow: All-Russian Research Institute of Hydraulic Engineering and Land Reclamation named after A.N. Kostyakov, 2022. 152 p. ISBN 978-5-907464-34-6. EDN BVVIEO.

7. Certificate of state registration of computer program No. 2025682023 Russian Federation. Intelligent web agent for database analysis: declared 29.07.2025; published 20.08.2025 / A.V. Matveev; applicant Federal State Budgetary Scientific Institution «Federal Scientific Center for Hydraulic Engineering and Land Reclamation named after A.N. Kostyakov». EDN FUDKJJ.

Матвеев Андрей Валерьевич, канд. техн. наук, вед. науч. сотрудник, andrey@vniigim.ru (ФНЦ «ВНИИГиМ имени А.Н. Костякова», Россия, г. Москва).

УДК 631.674.6.

DOI: 10.32962/0235-2524-2025-5-28-32

КАПЕЛЬНОЕ ОРОШЕНИЕ ЦВЕТОЧНЫХ РАСТЕНИЙ НА СКЛОНАХ ОВРАГА

М.И. ГОЛУБЕНКО, С.П. УМНОВ, Ю.А. МАЖАЙСКИЙ

Ключевые слова: капельное орошение, поливная вода, полив цветочных растений.

Keywords: drip irrigation, irrigation water, watering of flower plants.

Аннотация. Обзор литературных источников позволил установить, что в Российской Федерации проведение капельного орошения по использованию оросительной воды считается наиболее водосберегающим и экологически безопасным способом орошения, а по своей технической реализации наиболее автоматизированным.

В Российской Федерации имеется площадь капельного орошения до 150 тыс. га в 2021 г. Ежегодно происходит увеличение площади под капельным орошением и в 2025 г.

Для промышленного производства в Российской Федерации выпускаются ленты и трубки капельного полива в масштабах около 1 млрд п.м в год. Характеристики трубок позволяют подобрать необходимые параметры полива для любой культуры и почвенных особенностей. Выпускаются ленты и трубки капельного орошения диаметром 16 мм с тол-

щиной стенки от 0,15 до 1,2 мм с расстоянием между капельницами от 10 см и больше, расходом от 0,6 до 8 л/ч по давлению воды водовыпусками, позволяющими подобрать необходимую длину капельной линии в зависимости от величины участка.

При современном требовании экономии водных и энергетических ресурсов, сохранение и повышение плодородия почв, защиты растений и агроландшафтов в целом нашли подтверждение конкретно при выращивании растений «Патриарший сад» в центральной части города Владимир. На склонах оврага созданы террасы в сложных условиях рельефа, где размещается весь комплекс агроландшафта, в частности, полив цветочных растений, яблонь, вишни и других произрастающих растений, что делает сад завершенным для отдыха людей города со своим дизайном. Кроме того, обеспечивает зону отдыха в летний жаркий период на данном объекте.

Abstract. A review of the literature sources allowed us to establish that in the Russian Federation, drip irrigation for the use of irrigation water is