

2. Лазарев, Н. Н. Накопление валовой энергии бобово-злаковыми агрофитоценозами / Н. Н. Лазарев, С. М. Авдеев, Л. Ю. Демина // Доклады ТСХА. Том Выпуск 279, Часть 1. – Москва : Российский государственный аграрный университет, 2007. – С. 374–377. – EDN KHRDIO

УДК 631.171:004.8:634.1

**А.И. Кутырёв**, канд. техн. наук

*Федеральное государственное бюджетное научное учреждение  
«Федеральный научный агроинженерный центр ВИМ», г. Москва  
e-mail: alexeykutyrev@gmail.com*

## **МУЛЬТИМОДАЛЬНАЯ АГЕНТНАЯ ПЛАТФОРМА ДЛЯ ЦИФРОВОГО САДОВОДСТВА: ИНТЕГРАЦИЯ КОМПЬЮТЕРНОГО ЗРЕНИЯ, БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ И ИНТЕРНЕТА ВЕЩЕЙ**

**Ключевые слова:** агентный искусственный интеллект, большие языковые модели, RAG, компьютерное зрение, YOLO, глубокое обучение, интернет вещей.

**Keywords:** agentic artificial intelligence, large language models, RAG, computer vision, YOLO, deep learning, Internet of Things.

**Аннотация.** В статье представлена разработка мультимодальной агентной платформы для поддержки принятия решений в садоводстве, объединяющей методы компьютерного зрения (YOLO), большие языковые модели с механизмом дополненной генерации (RAG) и средства мониторинга данных с датчиков (MQTT, HTTP). Платформа предоставляет два интерфейса: Telegram-бот для оперативной работы в саду и веб-приложение для аналитики. Ядро системы включает локально развёрнутые модели DeepSeek-R1 (8B) и Qwen2.5-VL-7B-Instruct, векторную базу знаний ChromaDB, модули детекции и сегментации изображений, а также инструменты сравнительного тестирования и дообучения (QLoRA). Экспериментальная оценка показала, что применение RAG повысило релевантность ответов (86,2% предпочтений экспертов), система мониторинга обеспечивает приём до 500 записей/сек с задержкой оповещений 2 секунды, а реализованный алгоритм дообучения (QLoRA) позволяет адаптировать работу моделей под конкретные агротехнические задачи менее чем за 60 шагов обучения. Разработанная платформа может служить основой для автоматизированных рабочих мест агронома, систем дистанционной диагностики и раннего оповещения.

**Summary:** The article presents the development of a multimodal agent platform for decision support in horticulture, combining computer vision meth-

ods (YOLO), large language models with retrieval-augmented generation (RAG), and data monitoring tools from sensors (MQTT, HTTP). The platform provides two interfaces: a Telegram bot for operational work in the orchard and a web application for analytics. The system core includes locally deployed DeepSeek-R1 (8B) and Qwen2.5-VL-7B-Instruct models, a vector knowledge base ChromaDB, image detection and segmentation modules, as well as tools for comparative testing and fine-tuning (QLoRA). Experimental evaluation showed that the use of RAG increased answer relevance (86.2% expert preference), the monitoring system supports up to 500 records/sec with an alert delay of 2 seconds, and the implemented fine-tuning algorithm (QLoRA) allows adapting the models' operation to specific agronomic tasks in less than 60 training steps. The developed platform can serve as a basis for automated agronomist workstations, remote diagnostic systems, and early warning systems.

Современное промышленное садоводство находится на этапе активной цифровизации: внедряются автоматизированные метеостанции, стационарные и мобильные датчики мониторинга состояния растений и почвы, IoT-платформы, беспилотные летательные аппараты и роботизированные платформы [1,2]. В результате формируются большие объёмы разнородных данных, требующих эффективных методов обработки. Интерпретация этих данных в готовые управленческие решения остаётся сложным многоэтапным процессом, требующим высокой квалификации и использования различных программных инструментов [3]. Для автоматизации обработки визуальной информации применяются свёрточные нейронные сети, в частности семейства YOLO, позволяющие вести подсчёт плодов, цветков, завязей и диагностировать болезни [4-6]. В свою очередь, большие языковые модели (LLM) и их мультимодальные расширения (VLM) дают возможность получать развёрнутые текстовые заключения по результатам анализа изображений [7,8]. Технология генерации с дополненным поиском (RAG) позволяет дополнять знания модели внешними проверенными источниками, что важно для узкоспециализированных областей [9]. Системы мониторинга на основе MQTT и анализа потоковых данных широко применяются в «умном сельском хозяйстве» [10]. Несмотря на активное развитие отдельных направлений, комплексные решения, объединяющие обработку естественного языка, компьютерное зрение, потоковый анализ данных и предоставляющие различные интерфейсы доступа (веб и мессенджеры), остаются недостаточно разработанными. Целью исследований является создание мультимодальной агентной платформы для цифрового садоводства, интегрирующей методы компьютерного зрения, большие языковые модели с механизмом RAG, средства мониторинга IoT-устройств и предоставляющей доступ как через веб-интерфейс, так и через Telegram-бота.

Разработанная платформа построена по модульному принципу на языке Python (рисунок 1). Ядро системы представляет собой серверное приложение на FastAPI, которое объединяет все вычислительные модули и предоставляет единый REST API. Поверх ядра реализованы два клиентских интерфейса: Telegram-бот (библиотека python-telegram-bot) для работы в саду и веб-приложение на Gradio для аналитики. Такая архитектура обеспечивает удобство использования, позволяя агроному, находясь непосредственно в саду, оперативно загрузить фото через бот, а аналитику – работать с дашбордами и проводить эксперименты через веб-интерфейс.

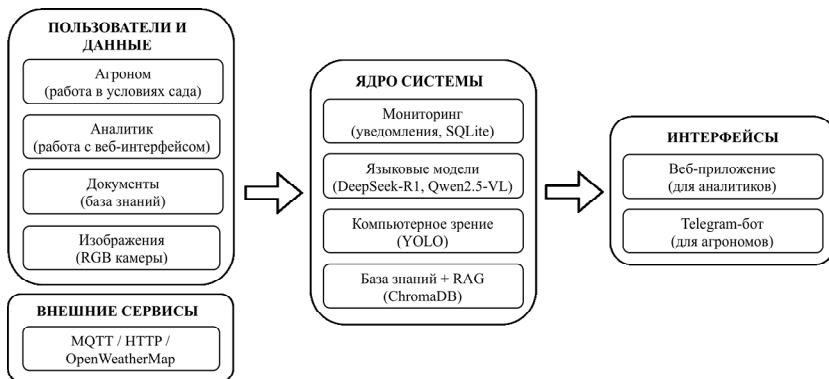


Рисунок 1. Архитектура мультимодальной агентной платформы

Основой для решения задач компьютерного зрения служат модели семейства YOLO (библиотека Ultralytics). Платформа поддерживает распознавание объектов, сегментацию экземпляров, семантическую сегментацию, классификацию и определение ключевых точек (например, скелет кроны деревьев). Используются модели YOLOv8-v12, дообученные на специализированной выборке из более чем 30000 изображений промышленных садов. Пользователь может настраивать пороги уверенности (conf) и пересечения (iou) непосредственно в интерфейсе. В Telegram-боте реализована возможность выбора из более чем 15 специализированных классов, включая плоды, цветки, завязи, признаки болезней и др.

Платформа интегрирует два типа языковых моделей. Для мультимодального анализа изображений (распознавание и интерпретация визуальной информации) используется локально развернутая модель Qwen2.5-VL-7B-Instruct (LM Studio), которая обрабатывает размеченные YOLO изображения и формирует текстовые заключения. Для консультирования на естественном языке, работы с базой знаний и ответов на профессиональные вопросы применяется модель DeepSeek-R1 (версия deepseek/deepseek-r1-0528-qwen3-8b), также загруженная локально в 4-битном формате.

Модуль базы знаний и RAG реализован на основе векторной базы данных ChromaDB. Для оценки смысловой близости используется косинусная метрика сравнения эмбеддингов, создаваемых моделью SentenceTransformer intfloat/multilingual-e5-small (размерность 384). Пользователь может загружать документы в форматах PDF, DOCX, TXT, XLSX, CSV, HTML. Система извлекает текст, сегментирует его на фрагменты (по умолчанию 500 символов с перекрытием 50), вычисляет эмбеддинги и сохраняет в индексе FAISS. При поступлении запроса извлекаются 3-5 наиболее релевантных фрагментов, которые вместе с вопросом передаются в языковую модель. Модуль мониторинга и уведомлений обеспечивает сбор данных от трёх типов источников, включая ручной ввод через интерфейс, подписка на MQTT-топики (библиотека paho-mqtt), а также HTTP-эндпоинт /api/monitoring, принимающий POST-запросы с массивом записей. Каждая запись содержит временную метку, название метрики, числовое значение и идентификатор источника. Данные сохраняются в SQLite. Пользователь может создавать правила уведомлений. При срабатывании правила формируется оповещение, отправляемое в Telegram через HTTP-запросы. Интеграция с OpenWeatherMap API позволяет дополнять данные погодной информацией.

Все взаимодействия пользователя с системой (вопросы, ответы, режим использования RAG, оценки релевантности) сохраняются в базу данных. Для сравнения эффективности различных режимов реализован инструмент сравнительного тестирования. Пользователь получает два варианта ответа (с RAG и без) и голосует за лучший. Накопленные диалоги могут быть экспортированы в формате JSONL для последующего дообучения. Реализована возможность дообучения языковой модели методом QLoRA с использованием библиотеки Unsloth. Процесс начинается со сбора датасета из логов взаимодействий (пары «вопрос – ответ»). Базовая модель загружается с 4-битным квантованием, применяются адаптеры LoRA. Гиперпараметры: batch size = 2, gradient accumulation = 4, learning rate = 0,0002. Обучение занимает менее 60 шагов и требует не более 5 ГБ видеопамати. В качестве интерфейса для агрономов использован Telegram-бот. Он поддерживает загрузку изображений и видео, выбор задачи компьютерного зрения (детекция, сегментация, классификация), настройку параметров моделей, диалог с VLM и RAG, а также сбор структурированных датасетов с метаданными (геолокация, погодные условия). Все результаты возвращаются в виде размеченных изображений и текстовых заключений. Веб-интерфейс на Gradio предоставляет расширенные возможности для аналитиков и исследователей: семь основных вкладок (AI-агент с различными режимами чата, мониторинг с графиками, анализ изображений с настройкой YOLO, управление уведомлениями, работа с датасетами и сравнительное тестирование, управление базой знаний, администрирование и резервное копирование) (рисунок 2).

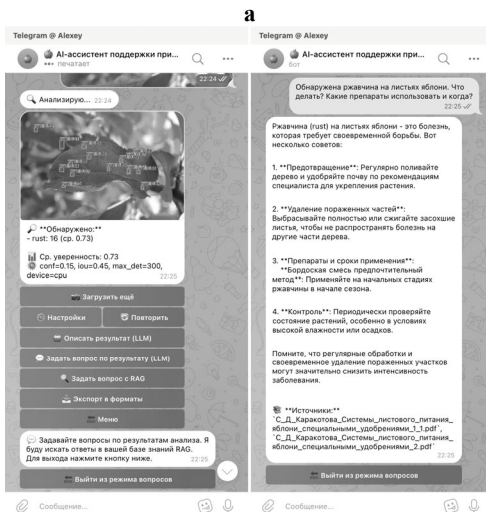
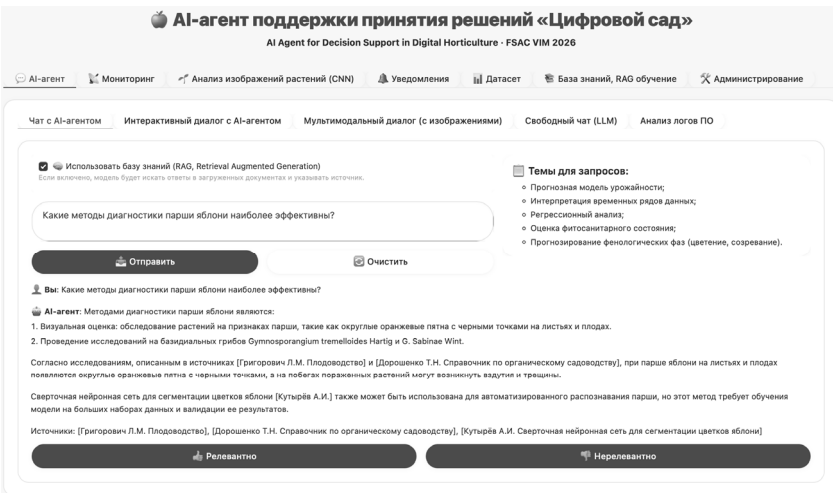


Рисунок 2. Мультимодальная агентная платформа для поддержки принятия решений в садоводстве: а – веб-интерфейс платформы, б – Telegram-бот

Эксперименты проводились на изображениях и данных мониторинга, полученных в 2024-2025 гг. в промышленных садах Курской и Московской областей. Подготовлено 70 вопросов по садоводству (интерпретация данных мониторинга, прогнозирование фенологических фаз, фитосанитарный контроль). В базу знаний загружены справочники и научные публика-

ции. Сравнение двух версий ответов (с RAG и без) с участием 10 экспертов показало, что в 86,2% случаев предпочтение отдано ответам с использованием базы знаний. Согласованность экспертов по капле Коэна составила 0,72. Время ответа увеличивалось в среднем на 0,8 секунды, что не нарушает интерактивность и подтверждает применимость алгоритма в реальном времени. Интеграция моделей YOLO в единый интерфейс обработки запросов Telegram-бота и веб-интерфейса показала стабильную работу при параллельной обработке потока изображений. Время инференса на GPU NVIDIA GeForce RTX 2080 Ti составило ~0,05 с на кадр, что обеспечивает возможность оперативного ответа непосредственно в саду. Относительная ошибка автоматизированного подсчёта плодов, выполняемого в рамках единого агентного цикла («изображение → детекция → аналитическое заключение»), не превысила 7% по сравнению с ручным экспертным учётом, а коэффициент детерминации между автоматическим и экспертным подсчётом составил  $R^2 = 0,94$ . Оценка мониторинга и событийной интеграции. Эмуляция работы датчиков с передачей значений температуры через MQTT показала, что при настройке правил (температура ниже 2°C – риск заморозков, выше 30°C – необходимость полива) уведомления в Telegram поступают со средней задержкой 2 секунды. Данные результаты подтверждают пригодность платформы для работы в режиме реального времени с промышленными IoT-устройствами.

Оценка возможностей адаптации (дообучения). После дообучения на 400 диалогах с положительными оценками (метод QLoRA) языковая модель стала давать более развёрнутые рекомендации с указанием источников. Например, на вопрос о борьбе с паршой яблони обученная модель предложила конкретные препараты и схему обработки, ссылаясь на загруженные в RAG методические материалы, тогда как исходная давала лишь общую рекомендацию. Реализованный модуль дообучения (менее 60 шагов, <5 ГБ VRAM) позволяет оперативно адаптировать поведение агентной платформы под нужды конкретного хозяйства без необходимости полного переобучения. В ходе опытно-промышленного тестирования с участием пяти агрономов зафиксировано сокращение времени на формирование заключения (анализ фото + данные датчиков + поиск в базе знаний) с 15-20 минут (при традиционном способе) до 2-3 минут при использовании платформы. Система выдержала пиковую нагрузку 50 параллельных запросов к Telegram-боту с сохранением времени ответа менее 5 секунд для 95% запросов, что подтверждает масштабируемость архитектуры на базе FastAPI.

В результате проведенных исследований разработана мультимодальная агентная платформа для цифрового садоводства, объединяющая компьютерное зрение (YOLO), большие языковые модели (DeepSeek-R1, Qwen2.5-VL) с механизмом RAG, средства мониторинга IoT-устройств (MQTT, HTTP) и предоставляющая доступ через Telegram-бота и веб-интерфейс. Экспериментально подтверждена эффективность предложен-

ной мультимодальной агентной платформы, применение RAG повысило релевантность ответов (86,2% предпочтений экспертов), система мониторинга обеспечивает приём до 500 записей/сек с задержкой оповещений 2 секунды, а реализованные инструменты сравнительного тестирования и дообучения (QLoRA) позволяют адаптировать платформу под условия конкретного хозяйства, накапливать структурированные датасеты и повышать качество рекомендаций. Практическое применение платформы позволило сократить время принятия решений агрономом в 5–7 раз, обеспечить масштабируемость до 50 параллельных подключений и организовать автоматическое пополнение обучающих выборок. Агентная платформа может служить основой для создания автоматизированных рабочих мест агронома, систем дистанционной диагностики болезней, раннего оповещения о неблагоприятных событиях (заморозки, засуха, вредители) и модулей прогнозирования урожайности.

#### Список использованной литературы

1. Poenaru M.M., Manta L.F., Gherțescu C., Manta A.G. Shaping the Future of Horticulture: Innovative Technologies, Artificial Intelligence, and Robotic Automation Through a Bibliometric Lens // *Horticulturae*. 2025. № 11(5). 449.

2. Kutyrev A., Andriyanov N., Khort D., Smirnov I., Zubina V. Adaptive CNN Ensemble for Apple Detection: Enabling Sustainable Monitoring Orchard // *AgriEngineering*. 2025. № 7(11). 369. <https://doi.org/10.3390/agriengineering7110369>

3. Kalaany CMA, Kimaita HN, Abdelmoneim AA, Khadra R, Derardja B, Dragonetti G. The Potential of Low-Cost IoT-Enabled Agrometeorological Stations: A Systematic Review // *Sensors*. 2025. № 25(19). 6020.

4. Кутырёв А.И. Распознавание и классификация болезней листьев яблони на основе анализа их изображений моделями сверточных нейронных сетей (CNN) // *Вестник Ульяновской государственной сельскохозяйственной академии*. 2023. № 3 (63). С. 215–223.

5. Кутырёв А.И., Смирнов И.Г., Андриянов Н.А. Сравнительный анализ моделей нейронных сетей для распознавания плодов яблони на кроне дерева // *Садоводство и виноградарство*. 2023. № 5. С. 56–63.

6. Кутырёв А.И. Сверточная нейронная сеть для сегментации цветков яблони на изображениях // *Аграрная наука Евро-Северо-Востока*. 2024. Т. 25. № 5. С. 949-961.

7. Yu P., Lin B. A Framework for Agricultural Intelligent Analysis Based on a Visual Language Large Model. *Applied Sciences*. 2024. № 14(18). 8350.

8. Yu G-H., Anh L.H., Vu D.T., Lee J., Rahman Z.U., Lee H-Z., Jo J-A., Kim J-Y. VL-PAW: A Vision-Language Dataset for Pear, Apple and Weed // *Electronics*. 2025. № 14(10). 2087.

9. Bai B, Meng X, Zhao C. Research on Sem-RAG: A Corn Planting Knowledge Question-Answering Algorithm Based on Fine-Grained Semantic Information Retrieval Enhancement. *Applied Sciences*. 2025; 15(19):10850.

10. Has M., Kreković D., Kušek M., Podnar Žarko I. Efficient Data Management in Agricultural IoT: Compression, Security, and MQTT Protocol Analysis // *Sensors*. 2024. № 24(11). 3517.