УДК 336.719

ИСПОЛЬЗОВАНИЕ АГЕНТОВ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РЕШЕНИЯ АНАЛИТИЧЕСКИХ ЗАДАЧ

Андрианов В. А.[1]

магистрант,

Финансовый университет при Правительстве Российской Федерации,

Москва, Россия

Аннотация. В статье рассматриваются архитектура и ключевые компоненты ИИ-агентов: языковая модель («мозг»), системы восприятия информации и инструменты для выполнения действий. Особое внимание уделяется применению таких агентов в аналитических задачах, включая методы RAG и text-to-sql, позволяющие интегрироваться с корпоративными данными. Анализируются преимущества агентного подхода перед традиционными BI-системами, а также текущие ограничения внедрения, связанные со сложностью технологии и нестабильностью работы моделей.

Ключевые слова: ИИ-агенты, большие языковые модели, аналитические задачи, RAG, text-to-sql, искусственный интеллект (ИИ), автоматизация.

USING ARTIFICIAL INTELLIGENCE-BASED AGENTS TO SOLVE ANALYTICAL PROBLEMS

Andrianov V. A.[2]

Master's Student,

[1] **Научный руководитель: Громова Алла Александровна**, к.э.н., доцент, кафедра «Финансовые технологии» Финансового факультета ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации», Россия, г. Москва

[2] **Scientific supervisor: Alla Gromova**, PhD in Economics, Associate Professor, Department of Financial Technologies, Faculty of Finance, Federal State Budgetary Educational Institution of Higher Education, Financial University under the Government of the Russian Federation, Moscow, Russia

Financial University under the Government of the Russian Federation,
Moscow, Russia

Abstract. The article discusses the architecture and key components of AI agents: the language model ("brain"), information perception systems, and tools for performing actions. Special attention is paid to the use of such agents in analytical tasks, including RAG and text-to-sql methods that allow integration with corporate data. The advantages of the agent-based approach over traditional BI systems are analyzed, as well as current implementation limitations related to the complexity of the technology and the instability of the models.
Keywords: AI agents, large language models (LLM), analytical tasks, RAG, text-to-sql, artificial intelligence (AI), automation.

В последние годы самой прорывной технологией стал искусственный интеллект (далее — ИИ). Началом нового витка развития ИИ стал 2022 год, когда компания OpenAI выпустила свой чат-бот ChatGPT. Уже в 2024 году совокупный объем инвестиций в ИИ превысил 1 триллион долларов [8], а вложения в дата-центры, дающие необходимые вычислительные мощности, в 2025 году превзошли вложению в нефтяную отрасль [1].

За эти 3 года произошел не только рост вложений в ИИ. Стали более разнообразными направления использования больших моделей на основе искусственного интеллекта. Если изначально все компании-пионеры отрасли стремились к созданию LLM, больших языковых моделей, чтобы в будущем продавать доступ к ней по подписочной системе, то к 2025 году акценты сместились на создание ИИ-агентов.

ИИ-агент — это программный объект, который автономно выполняет задачи, анализируя внешний мир и используя специальные инструменты. В его

основе лежит большая языковая модель, например, в ChatGPT, однако агент не только выдает ответ — он может на основе найденного сделать конкретное действие. Помимо языковой модели, агент использует внешние инструменты и функции (рассуждение, планирование). Это позволяет совершать многошаговые действия, которые приводят к достижению конкретного результата (рис. 1).

Рисунок 1 - Основные компоненты ИИ-агента [3]

Другими словами, у любого ИИ-агента есть мозг (большая языковая модель), восприятие (то, что заменяет в агенте органы чувств человека) и выполнение действий (с помощью инструментов). Подробнее рассмотрим каждый элемент.

Мозг ИИ-агента — это, в своей основе, языковая модель. Например, Claude Sonnet от Anthropic, GPT от OpenAI, Gemini от Google или Alice AI от Яндекса. Для эффективной работы эта модель должна обладать несколькими характеристиками (рис. 2).

- Взаимодействие на естественном языке. То есть мы обращаемся к модели на понятном нам языке, например, русском или английском, без использования специальных символов или искусственных языков.

- Обучение на различных данных. Модель должна обладать как общими знаниями, так и специализированными (программирование, медицина). Но самое главное, для достижения результата модель должна быть «здравомыслящей», то есть обладать базовым пониманием мирового устройства. Нехватка такого понимания приводит к галлюцинациям модели, и как следствие, к неверным действиям агента.
- Память, которая включает последовательность наблюдений, мыслей и действий агента. Проблема памяти характерна для всех моделей, потому что она ограниченна из-за сложностей хранения.
- Планирование и рассуждение. Рассуждение, то есть процесс создания утверждений в логической последовательности при определенных обстоятельствах повышает эффективность работы агента. То же касается и планирования. Планирование обычно состоит из декомпозиции и последующей оценки. Ввиду технической сложности планирование и рассуждение иногда могут вредить агенту, замедляя его работу и увеличивая вероятность галлюцинаций.

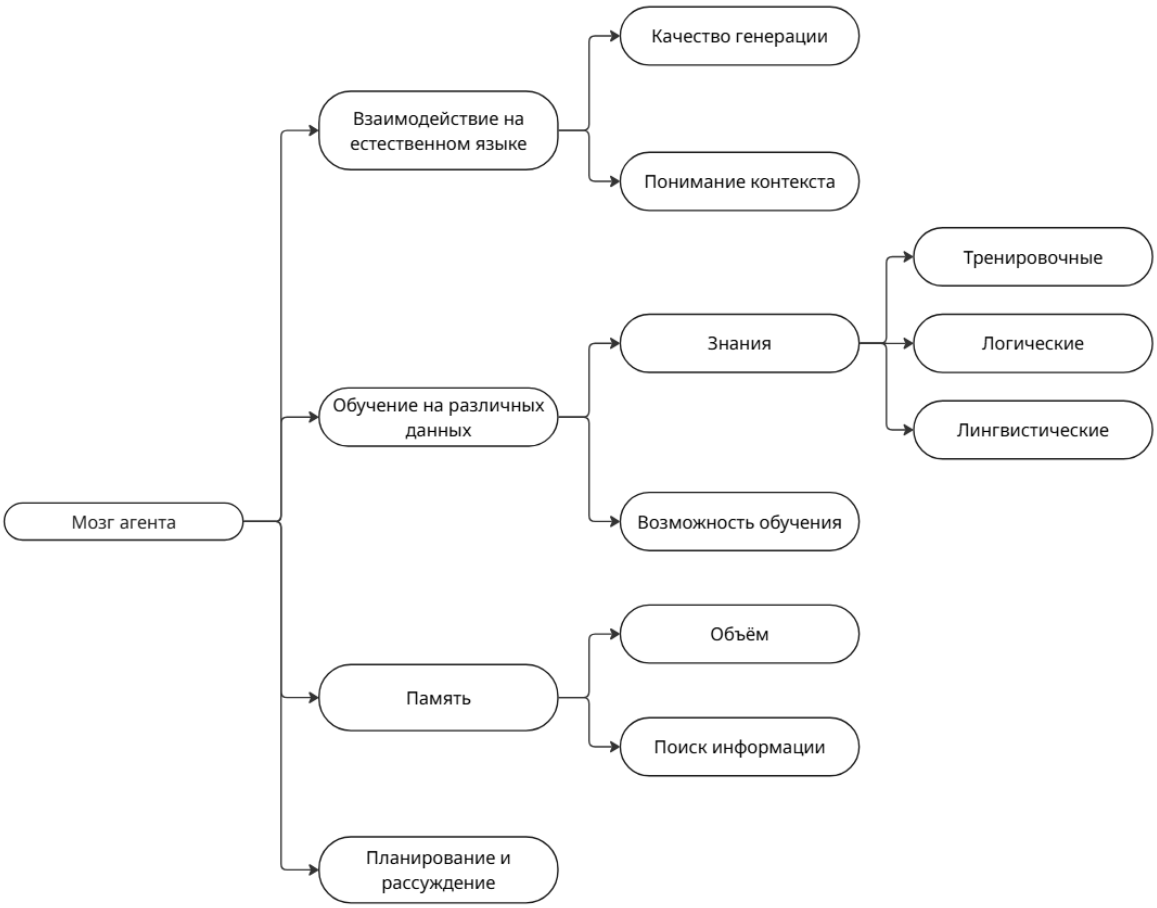

Рисунок 2 - Характеристики первого компонента ИИ-агента — мозга [фф]

Второй компонент ИИ-агента — это способность агента к восприятию. В качестве объекта восприятия может быть текстовая, визуальная, аудиальная и сенсорная информация.

Безусловно, в основном, на вход подается текстовая информация. На ней обучены языковые модели, с помощью текста часто общаются люди. При этом все равно возникают сложности с правильной интерпретацией человеческого контекста, потому что модель воспринимает текст без внутреннего подтекста, который у человека произрастает из здравого смысла и логики.

Визуальная информация воспринимается моделью тяжелее. Для упрощения работы с картинками и видео для модели создается текстовое описание, то есть, по сути, визуальная информация трансформируется в текстовую, что менее эффективно, особенно на больших объемах. Поэтому

активно развивается такая область, как компьютерное зрение. Применительно к видеоинформации используется то же компьютерное зрение, а видео просто делится на много картинок, следующих одна за одной.

Чтобы воспринимать аудиоинформацию, модель разбивает звуковой поток на части. Они обрабатываются, фильтруются и переводятся в текстовый формат.

Что касается сенсорной информации, она может включать в себя жесты, движение глаз, тела, изменение в мозге человека, изменение температуры — то есть все, что происходит во внешнем мире (рис. 3).

Очевидно, что на сегодняшний день большинство агентов работает с текстовой информацией, потому что она наиболее универсальная. Крупнейшие LLM-модели обучены именно на массиве текстовой информации в сочетании с визуальной.

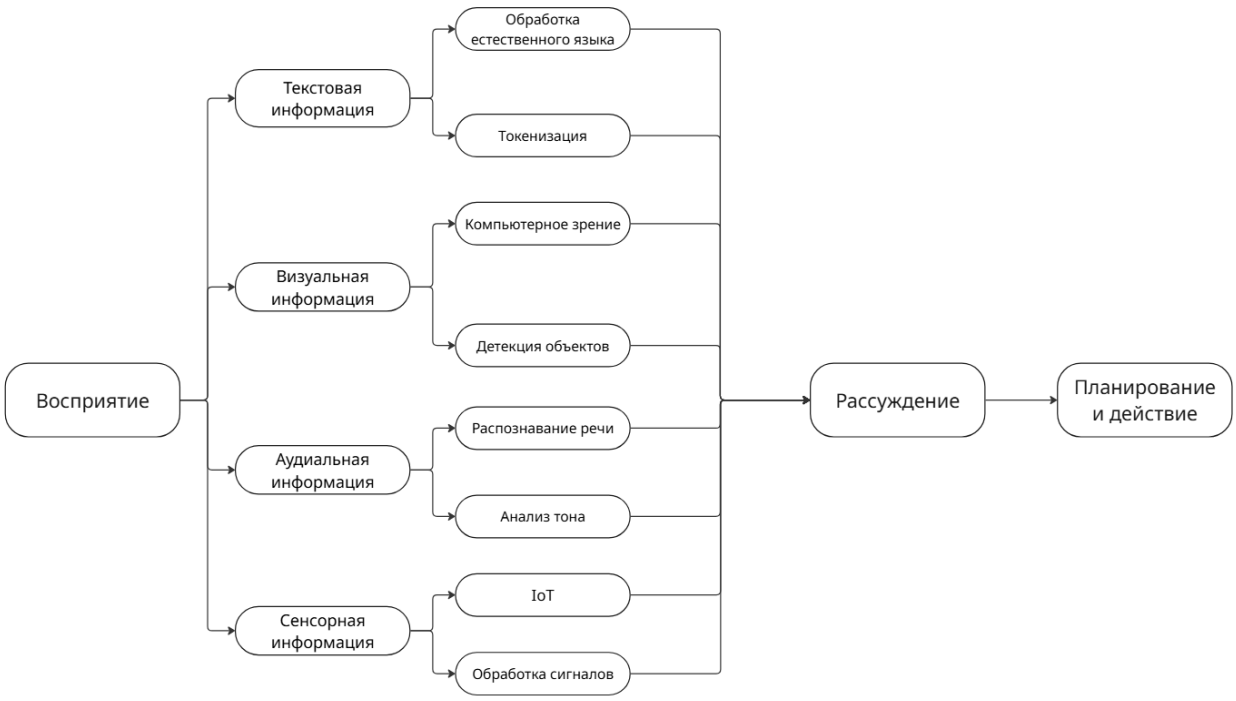

Рисунок 3 - Характеристики второго компонента ИИ-агента — восприятия [6]

Третий элемент — способность к действию и выдаче определенного результата. Это может быть результат в виде текста, как мы привыкли в чат-ботах, в форме графика, рисунка или в табличном виде. Так же агент может

обращаться к инструментам, после использования которых будет дан ответ пользователю. Еще одним возможным результатом было бы совершение действий роботом в реальном мире. Сейчас такие случаи единичны из-за сложности и дороговизны реализации (рис. 4).

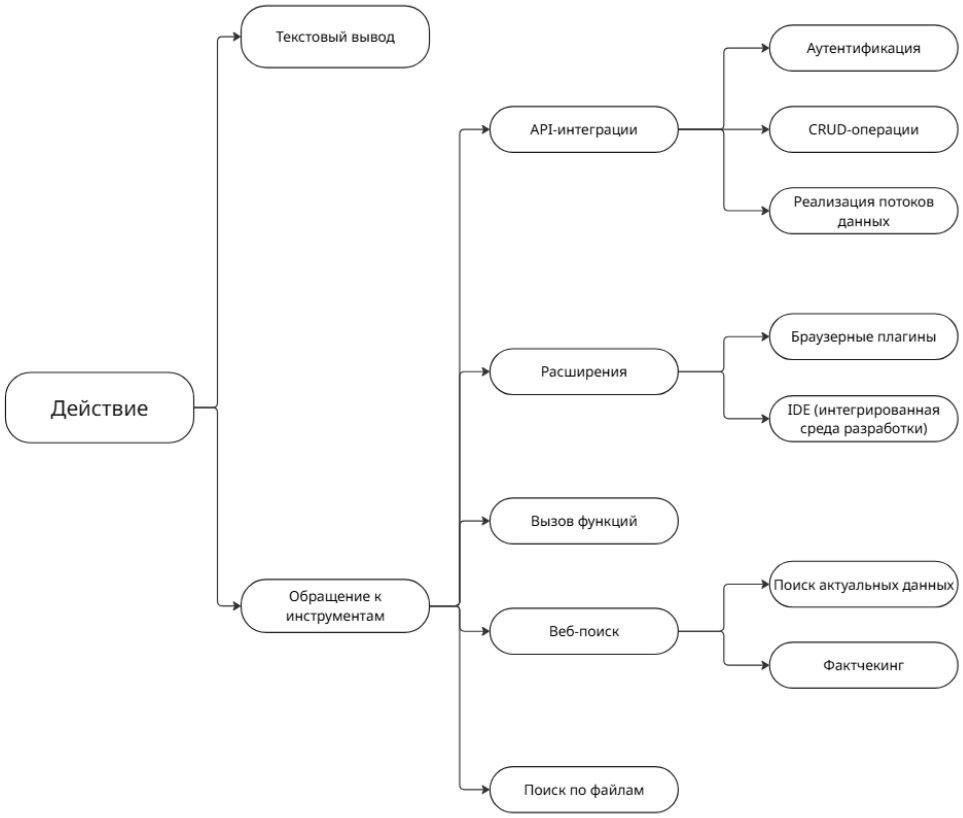

Рисунок 4 - Характеристики третьего компонента ИИ-агента — действия [7]

Одним из главных ресурсов, благодаря которым агент может взаимодействовать с внешним миром — это инструменты. Если обычные языковые модели не могут воспользоваться инструментами, так как они предобучены на ретроспективных данных, то агент может подключаться к инструментам и в реальном времени получать информацию из внешних источников.

Одним из основных способов подключения к внешнему источнику — это вызов API. Например, Google Flights API предоставляет данные об авиаперелетах. Для подключения агента к API-сервису используются

расширения — инструкции для агента, при каких запросах обращаться к API, а при каких нет.

Еще одним популярным инструментом являются хранилища данных. В них хранится дополнительная информация, связанная с определенной компанией или специфической деятельностью. То есть любая информация, которую не может знать базовая языковая модель в силу ограниченности данных, на которых она обучена.

Примерами такого хранилища данных могут быть текстовые файлы и данные с веб-сайтов. Сложность подключения агента к хранилищам заключается в том, что хранилища плохо структурированы. При этом все сырые данные перед использованием агентом могут быть преобразованы в векторную базу данных. Процесс подключения агента к внешним базам данным для выдачи ответа на их основе называется RAG (Retrieval-Augmented Generation, генерация, дополненная поиском). Самая важная часть — это подготовка данных [11].

Исходные документы любого формата (pdf, docx, html) разбиваются на небольшие смысловые части. Каждая часть обрабатывается моделью векторизации. Она преобразует части в векторы, которые находятся в определенной точке многомерного пространства. Части с похожими характеристиками будут находиться рядом в многомерном пространстве. Эти векторы вместе с исходными данными сохраняются в векторную базу данных (рис. 5).

Рисунок 5 - Этапы формирования ответа с помощью RAG [2]

После формирования запроса агент сможет преобразовать его с помощью модели векторизации в вектор, который направится в векторную базу данных. В базе данных будет осуществлен поиск наиболее близких векторов. На основе полученных векторов сформируется контекст. Агент найдет похожие векторы в базе данных и сможет выдать релевантный ответ.

Инструменты, подобные RAG, часто используют при создании ИИ-агентов для аналитических задач. Например, так агент может получать информацию из внутренних документов компании, и на основе ее и информации из интернета, выдавать более точные результаты.

Также распространяется использование text-to-sql — метода, при котором запрос на естественном языке переводится в запрос на специализированном для баз данных языке SQL. Этот запрос направляется в базы данных и после получения результатов переводится из табличного формата в наиболее удобный человеку (график, отдельное число или сводная таблица) (рис. 6).

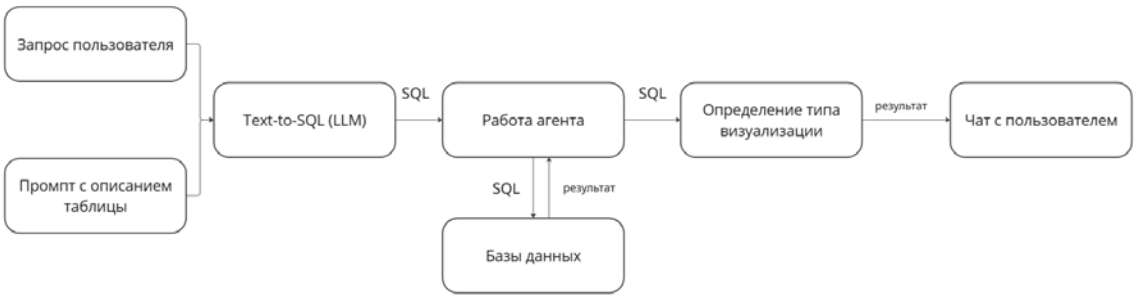

Рисунок 6 - Работа аналитического ИИ-агента с помощью метода text-to-sql [10]

Чтобы LLM понимал, как вытянуть из витрины необходимые данные, нужно точно описать в текстовом виде, из чего состоит таблица. То есть нужно определить поля, тип данных, пояснения к столбцам, примеры значений. Дальше происходит настройка агента под разные возможные ситуации. В сложных случаях возможен переспрос запроса.

Очевидно, что для получения данных из таблицы запрос должен содержать релевантную этой таблице информацию. Подобная настройка происходит при помощи написания промптов. Например, «Проверь, насколько запрос соответствует данным, которые есть в базе данных» (дальше перечисляются критерии определения таких данных). Или «Ты — аналитик данных с многолетним стажем. Ты получаешь запрос, который ты должен классифицировать как релевантный доступной тебе базе данных или не относящийся к ней». Или «Проанализируй запрос, соотнеси его со схемой базы данных и в случае, если информации недостаточно, задай уточняющий вопрос».

После того, как агент определил запрос как подходящий, он генерирует SQL-код, учитывая подсказки по базе данных, схему базы данных, примеры расчета метрик — то есть всю информацию, которая поможет агенту корректно написать запрос к базе данных. Написанный агентом код может быть некорректным, что приведет к выдаче неверных результатов на выходе, поэтому этот код необходимо валидировать.

Существует несколько подходов к валидации кода, например, Spider, Bird, CoSQL [4]. Каждый состоит из множества тестовых наборов данных, а качество модели определяется долей совпадений эталонного SQL-кода и кода от агента. Если быть точнее, сравнивается не сам код, а те столбцы и поля, что он выдает. Несмотря на многоступенчатую оценку сгенерированных ответов, агент все еще подвержен ошибкам из-за галлюцинаций, синтаксических ошибок (есть много разных диалектов SQL, в которых агент может путаться) и логических ошибок (несоответствие логики языка SQL).

На выходе пользователь получает ответ в нескольких возможных форматах, которые определяются самим агентом на основе запроса. Если запрос направлен на получение одной или нескольких конкретных показателей, может выдаваться текст. Если необходима динамика или структура — линейный график или столбчатая диаграмма. Если модель не может определить формат, то данные зачастую выдаются в табличном формате.

Внедрение ИИ-агентов пока ограниченно из-за сложности и новизны технологии, а также из-за галлюционирования моделей, что ведет к нестабильности работы в промышленных масштабах. Вместе с тем, агенты обладают значимыми преимуществами по сравнению с другими аналитическими инструментами уже сейчас.

Например, в отличие от BI-систем с фиксированной структурой и моделью данных, агент обращается напрямую к схеме базы данных. Поэтому у него есть возможность строить запросы любой сложности, включая соединения множества таблиц, подзапросы и агрегации. Единственным ограничением тут является только способность агента строить сложные запросы хорошего качества [5].

Также агент может не просто выполнить одиночный запрос, но и, получив результат, задать уточняющие вопросы, предложить гипотезы и выявить

аномалии. Например, он может справиться с вопросом по типу «С чем связано падение метрики?» Сначала агент сделает запрос к базе данным, потом проведет анализ различных факторов, влияющих на эту метрику (согласно документации или после поиска в интернете), а в конце выведет не только количественную оценку, но и возможные причины падения в формате, удобном для человека.

Очевидным плюсом агентов является то, что пользователь может не просто написать запрос на естественном языке, но и вести дальнейший диалог, уточняя или задавая новые вопросы. При этом в случае недоверия к полученным данным пользователь сможет спросить у агента, откуда эти данные были получены, какая использовалась формула расчета и насколько информация актуальна.

Как только будет решена проблема стабильности и надежности работы ИИ-агентов, их повсеместное применение станет практически неизбежным. Ведь агент обладает как базовыми знаниями, благодаря наличию большой языковой модели, так и специфическими навыками, благодаря возможности подключению к внутренним данным компаний.

Библиографический список:

1. Новая нефть: почему инвестиции в ИИ стали главным мировым трендом. РБК Новости. URL: https://companies.rbc.ru/news/NqsnabuvNj/novaya-neft-pochemu-investitsii-v-ii-stali-glavnyim-mirovyim-trendom/ (дата обращения: 27.03.2026).

2. Huang H. Democratizing LLM Efficiency: From Hyperscale Optimizations to Universal Deployability. — 2025. URL: https://arxiv.org/abs/2511.20662 (дата обращения: 27.03.2026).

3. Liu X. AgentBench: Evaluating LLMs as Agents. — 2023. URL: http://arxiv.org/abs/2308.03688 (дата обращения: 27.03.2026).

4. Masterman T. The landscape of emerging ai agent architectures for reasoning, planning, and tool calling: a survey. — 2024. URL: https://arxiv.org/pdf/2404.11584 (дата обращения: 27.03.2026).

5. Na L. From LLM to Conversational Agent: A Memory Enhanced Architecture with Fine-Tuning of Large Language Models. — 2024. URL: http://arxiv.org/abs/2401.02777 (дата обращения: 27.03.2026).

6. Shinn N. Reflexion: Language Agents with Verbal Reinforcement Learning. — 2023. URL: http://arxiv.org/abs/2303.11366 (дата обращения: 27.03.2026).

7. Shunyu Y. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. — 2023. URL: http://arxiv.org/abs/2305.10601 (дата обращения: 27.03.2026).

8. Text2SQL в аналитике: как мы научили ИИ понимать бизнес-запросы без посредников URL: https://habr.com/ru/companies/X5Tech/articles/949694/ (дата обращения: 27.03.2026).

9. The 2025 AI Index Report. Stanford HAI. URL: https://hai.stanford.edu/ai-index/2025-ai-index-report/economy (дата обращения: 27.03.2026).

10. Xu H. Understanding the planning of LLM agents: A survey. — 2024. URL: https://arxiv.org/abs/2402.02716 (дата обращения: 27.03.2026).

11. Zhiheng X. The rise and potential of large language model based agents: a survey. — 2025. URL: https://arxiv.org/pdf/2309.07864 (дата обращения: 27.03.2026).