

25 ноября 2024

ИИ-аналитики, мультиагенты и генераторы кодов: спрос на ИИ-решения в российском бизнесе вырос на 40%

Объем проектов по работе с данными и искусственным интеллектом в России вырос на 40% в 2024 г. Главные направления - внедрение ИИ для общения и анализа, мультиагентный подход и внедрение генераторов кода в процесс разработки.

Об этом сообщило ООО "Рексофт" сегодня, 25 ноября. Эксперты компании отметили в 2024 г. 40%-ный рост заказов в областях, связанных с интеграцией искусственного интеллекта и работой с данными, по сравнению с 2023 г. Главной причиной роста, по их мнению, является активизация импортозамещения и переход многих компаний на модели, снижающие возможность вендор-лока.

Эксперты отметили рост популярности генеративного искусственного интеллекта в бизнес-процессах во всех отраслях экономики, от ретейла до промышленности. Среди наиболее востребованных технологий они назвали техники, дополняющие знания больших языковых моделей внутренними



данными, такие как RAG и Retrieval Augmented Generation. Такой подход полезен при разработке приложений, обрабатывающих частные данные или данные, полученные после обучения модели.

Компании России внедряют такие решения для обобщения и анализа документации, что позволяет в том числе создавать обучающие материалы для различных категорий специалистов. Эффективность этого подхода особенно заметна в компаниях от 1000+ сотрудников. Помимо разработки учебных материалов ИИ-ассистенты могут использоваться для оптимизации работы с какой-то специфичной нормативной документацией, например юридической.

Другой формой применения больших языковых моделей, по наблюдениям экспертов, становится так называемый мультиагентный подход, когда обученная на данных компании нейросеть берет на себя роль одного из сотрудников.

"Предварительная консультация с нейросетями помогает отсеивать наименее жизнеспособные идеи, а на обсуждение с командами выносить подготовленные аргументы. Подобная технология только начинает внедряться в процессы принятия решений, поэтому статистики эффективности не собрано, однако, по опыту применения, мультиагентный подход позволяет экономить один-два часа на задачу только на совещаниях

для каждого задействованного топ-менеджера", - сказано в сообщении "Рексофт".

Эксперты обратили внимание, что все больше компаний, как российских так и зарубежных, предлагают решения-ассистенты по генерации и анализу кода. Они предположили, что в ближайшем будущем начнется гонка производителей по внедрению подобных решений в полный цикл производства ПО: создание кода, верификацию, развертывание, тестирование, несмотря на то что сейчас эффективность таких решений не превышает 15%.

СПРОС РОЖДАЕТ ПРЕДЛОЖЕНИЕ

По словам руководителя дата-практики компании "Рексофт" Алексея Киселева, эксперты компании проанализировали количество отработанных человеко-часов на проектах в области данных за девять месяцев 2023 г. и за аналогичный период 2024 г., и разница между двумя периодами составила 40%. Главный драйвер спроса - крупные и средние компании, занятые в производстве, банковской сфере, страховании и ретейле.

Рост спроса отметил руководитель проектов по развитию AI-решений ООО "Директум" (Directum) Илья Петухов. Он сказал, что рост запросов на внедрение ИИ наблюдается во всех сферах развития технологии, от компьютерного зрения до больших языковых моделей.

"Статистика наших клиентов подтверждает тренд на выбор искусственного интеллекта. За последние три квартала мы фиксируем такую тенденцию: каждый второй клиент Directum (48%) выбирает вариант системы со встроенным ИИ. В 2023 г. этот показатель был на уровне 34-37%", - отметил он.

О значительном росте спроса рассказал основатель и исполнительный директор ИИ-платформы AutoFAQ Владислав Беляев. По его словам, только за осенний период 2024 г. спрос на услуги компании увеличился в три раза по сравнению с осенью 2023 г.

Этот рост интереса обусловлен несколькими факторами. Во-первых, бизнесы осознают ценность внедрения ИИ-решений для оптимизации процессов и повышения конкурентоспособности. Во-вторых, доступность инструментов ИИ и их популяризация через такие технологии, как языковые модели, стимулируют компании пробовать их в разных сферах деятельности. Однако мы также наблюдаем, что вместе с этим возникают завышенные ожидания от возможностей ИИ. Многие компании ожидают мгновенных результатов и универсальных решений, что часто приводит к неправильному пониманию требуемых инвестиций и времени на внедрение. Наша задача - не только предоставлять технологии, но и помогать клиентам правильно оценивать их возможности и внедрять ИИ с учетом реальных задач и ограничений.

О еще большем росте рассказала руководитель отдела анализа данных и машинного обучения АО "Наумен" Татьяна Зобнина. По ее словам, количество заказов на внедрение ИИ-решений выросло в два раза.

"Использование LLM как API, то есть когда большие языковые модели встраивают в приложения, продукты или сервисы, с начала года в Т-Банке выросло в девять раз. При этом использование LLM сотрудниками в ежедневной работе в компании с начала года выросло на 62%", - отметил руководитель LLM-платформы центр искусственного интеллекта АО "Т-Банк" Артур Самигуллин.

FROM RAGS TO...



Буквальный перевод аббревиатуры RAG - это "генерация с дополненной выборкой". Эта технология позволяет улучшить качество выходных данных языковой модели путем "подмешивания" в них соответствующей контексту запроса дополнительной информации. Ответ на вопрос, почему именно эти техники позволяют расширить области применения LLM, кроется в том, что, несмотря на то что языковые модели обучаются на огромных массивах данных, информация, которая используется для их обучения, статична. А значит, наборы знаний в этих моделях имеют явные ограничения в плане актуальности и глубины охвата конкретных областей. Именно подход с дополненной выборкой (RAG), путем обогащения запроса дополнительной актуальной информацией из внешних надежных источников, позволяет модели сформировать грамотный и верный ответ.

Ирина Абдуллаева,
младший научный сотрудник лаборатории Fusion Brain
института AIRI

По словам Артура Самигулина, RAG делает LLM доступным для бизнеса. Компании используют ее, потому что она не требует дообучения больших языковых моделей и помогает внедрять LLM поверх доменных данных без особых ML-навыков, сбора данных, алайнмента.

Ведущий менеджер продукта ООО "Контент ИИ" (Content AI) Антон Фролов обратил внимание, что создание LLM-модели требует доступа к большим объемам открытых данных. Это долгий и ресурсоемкий процесс, который невозможно повторить на данных из внутренних источников из-за высокой стоимости аппаратных ресурсов и необходимости учета прав доступа при обращении к документам. RAG-технология - выход из такой ситуации, так как сначала она ищет документы, релевантные запросу пользователя, а затем использует их и технологии LLM для формирования полноценного ответа.

"RAG может подсказывать и задавать направление генерации контента для LLM, а применение RAG вместе с поиском и LLM позволит существенно облегчить и ускорить поиск информации - например, во внутренних документах организации", - сказал Илья Петухов.

Алексей Киселев отметил, что в компании используется несколько корпоративных сервисов с RAG, которая улучшает работу языковых моделей. Среди них - ИИ-ассистент, который работает с корпоративной базой знаний, организует поиск внутри файлов, отвечая на вопрос на "естественном" языке, а не руководствуясь ключевыми словами. По его словам, этот сервис значительно сокращает время на поиск информации и подготовку коммерческого предложения.

МУЛЬТИАГЕНТ ПОКА НЕ 007

Мультиагентный подход - это когда несколько AI-агентов на основе LLM, каждый из которых умеет решать только назначенную ему подзадачу, вместе умеют решать некоторую исходную задачу пользователя. В идеале мультиагентный подход начинается с пользователя, который предоставляет задачу или запрос высокого уровня. Затем система разбивает задачу на более мелкие подзадачи и назначает их соответствующим специализированным агентам в зависимости от их ролей и возможностей. Существует много фреймворков для работы с мультиагентными системами, но на данный момент у специалистов в отрасли, которые занимаются решением реальных задач компаний, предприятий и государственных органов, данный подход не вызывает большого энтузиазма на практике, поскольку результаты скорее негативны. В силу негативности результатов говорить о ролях сотрудников, которые можно было бы поручить отдельным AI-агентам в таких системах, преждевременно.

Татьяна Зобнина

С ней согласна **руководитель группы Data Science департамента аналитических решений ООО "Корус Консалтинг ГК" Екатерина Торсукова**. По ее словам, мультиагентный подход сложно назвать состоятельным, и на данный момент ИИ не может заменить людей. Не в последнюю очередь потому, что все стратегические решения принимают

люди, а искусственный интеллект выполняет роль помощника.

Директор направления экспертизы платформы "Сфера" ООО "ТТ" Дмитрий Чувииков рассказал, что мультиагентный подход внедряется в работу компаний, особенно крупных. Там технология автоматизирует до 30% операций и берет на себя аналитику данных, планирование задач, распределение ресурсов и оценку рисков.

Согласно отчету Sequoia, 2023 г. стал годом активного внедрения копилотов. В 2024 г. ожидался постепенный переход к более сложным агентским системам, а 2025 г. прогнозируется как время активного развития и использования мультиагентов. В нашей работе мы активно исследуем и развиваем как агентские, так и мультиагентные системы, способные брать на себя функции сотрудников. Однако впереди еще много работы, чтобы эти технологии стали по-настоящему массовыми и эффективными.

Артур Самигуллин

Алексей Киселев сослался на опыт "Рексофт", согласно которому скорость проработки идеи проекта увеличивается на 7-10% при использовании мультиагентного подхода. Так, генеративные нейросети могут помочь специалисту посмотреть на идею с точки зрения разных специалистов и скорректировать ее согласно их видению.

"Помимо того, что этот подход может использовать непосредственно человек, часть корпоративных ИИ-сервисов также работают с ним. Мультиагентность здесь немного иного рода. У каждого из агентов внутри ИИ-сервиса задача, например, достать информацию из специализированного источника, распознать изображение таблицы, проверить работу другой модели и сделать общее заключение. Агенты взаимодействуют друг с другом, каждый в рамках поставленной задачи, и затем выдают готовый ответ", - сказал он.

ГЕНЕРАТОРЫ КОДОВ

По словам Ирины Абдуллаевой, генераторы кодов стали популярной практикой в среде разработчиков. Она отметила, что LLM-ассистенты для кодогенерации можно найти на любом языке программирования в любой интегрированной среде разработки (IDE).

"Подобные нейросети способны значительно упростить трудовую жизнь программисту, автоматизируя рутинные и повторяющиеся действия: генерация запросов для баз данных, API-запросы и стандартные функции. Ее можно применять для продумывания структуры проекта или отработки идей без затрат времени. Кроме того, генератор кода на базе LLM подходит для решения проблем с кодом и дебаггингом - он может исправить или объяснить ошибки в алгоритме", - рассказала она.

Некоторые компании разрабатывают собственные генераторы. Артур Самигуллин отметил, что сотрудники Т-Банка пользуются решением Nestor, чья задача упростить процесс программирования и снизить когнитивную нагрузку на разработчиков. В будущем они намерены сделать его универсальным помощником для ИТ-специалистов компании, а не только для программистов в IDE.

"Что касается генераторов кода, в "Рексофт" мы разработали и используем собственного помощника на базе LLM Deepseek-coder-v2. Эта современная модель показала наилучшие результаты в понимании контекста и написании качественного кода на основе наших требований. Например, для одного из клиентов мы выполнили перевод программного кода со старого технологического стека и БД Oracle на Java и PostgreSQL. С помощью ИИ мы сократили время на эту задачу в четыре-семь раз, сохранив высокое качество", - отметил Алексей Киселев.

Генераторы кода, такие как GitHub Copilot или ChatGPT, становятся все более популярными в разработке, однако важно понимать, что они выступают в роли помощников, а не полноценных заменителей разработчиков. Эти инструменты могут значительно ускорить процесс написания кода, особенно для стандартных или рутинных задач. Однако код, сгенерированный ИИ, может быть не всегда оптимальным, и в некоторых случаях его читаемость и качество могут оставлять желать лучшего. Программисты должны внимательно проверять и, при необходимости, дорабатывать предложенные решения.

Дмитрий Чувиков

Екатерина Торсукова более критична в оценке генераторов кода. Она назвала популярность таких решений хайпом и лайфхаком, которым пользуются студенты и джуны, но который не подходит для сложных задач и решений, так как ИИ пишет только несложный и неоптимизированный код, который нужно проверять и редактировать.

Наконец, по мнению генерального директора ООО "Базис" Давида Мартиросова, потенциал ИИ в разработке еще не раскрыт, так как "мы только входим в эру использования искусственного интеллекта". Однако он сослался на исследование GitHub, согласно которому более 97% разработчиков из

2000 опрошенных специалистов по всему миру хотя бы раз использовали ИИ в разработке.

