«БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ: ЭКСТРАВОЗМОЖНОСТИ ДЛЯ РОСТА БИЗНЕСА»

СОГЛАСНО ИССЛЕДОВАНИЮ КОМПАНИИ «ЯКОВ И ПАРТНЕРЫ», ЭКОНОМИЧЕСКИЙ ПОТЕНЦИАЛ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА (ИИ) В РОССИИ К 2028 ГОДУ СОСТАВИТ 22–36 ТРЛН РУБЛЕЙ, ПРИ ЭТОМ ЭФФЕКТ ГЕНЕРАТИВНОГО ИИ ОЦЕНИВАЕТСЯ В 1,3 ТРЛН, ЧТО БУДЕТ СОСТАВЛЯТЬ ОКОЛО 1% ВВП К 2028 ГОДУ. О ТОМ, КАКИЕ РОССИЙСКИЕ РЕШЕНИЯ УЖЕ ДОКАЗАЛИ ЭФФЕКТИВНОСТЬ И КАК С ПОМОЩЬЮ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ УМНОЖИТЬ РЕЗУЛЬТАТЫ, РАССКАЗАЛ ДМИТРИЙ ДЫРМОВСКИЙ, СЕО ГРУППЫ КОМПАНИЙ ЦРТ, КОТОРАЯ СТАЛА ПОБЕДИТЕЛЕМ В СПЕЦИАЛЬНОЙ НОМИНАЦИИ «ЦИФРОВОЙ ЭФФЕКТ» ПРЕМИИ «ТВЕРДЫЕ ЗНАКИ». ЕКАТЕРИНА ДЫННИКОВА

GUIDE: Как вы оцениваете развитие генеративного искусственного интеллекта?

ДМИТРИЙ ДЫРМОВСКИЙ: Первые версии моделей, которые сегодня фундаментально меняют действительность, появились более десяти лет назад, но глобальный прорыв в генеративном ИИ произошел в последние пару лет, когда появились такие решения, как ChatGPT. Сейчас генеративный AI проходит стадию роста, что подтверждают глобальные исследования. Так, к примеру, в последнем отчете аналитики McKinsey отметили, что уже 65% компаний используют GenAI хотя бы для одной бизнес-функции. Для сравнения, в прошлом году этот показатель был на уровне 33%. По нашим прогнозам, это направление в перспективе ближайших лет продемонстрирует еще более уверенный рост.

G: В чем его польза для бизнеса?

Д. Д.: Действительно полезным генеративный ИИ стал только в конце минувшего года, когда появились решения, где LLM (Large language model, большая языковая модель.— **G**) — это основа. Мы в ЦРТ фокусируемся на том, чтобы создать синергию речевых технологий и прогрессивных LLM, среди ведущих на российском рынке можно выделить нейросетевую модель GigaChat от «Сбера». Такая синергия уже сейчас помогает бизнесу: способствует росту выручки, снижению издержек, улучшению клиентского опыта, позитивно влияет на производственные процессы. В перспективе все больше речевых технологий будет объединяться в большие мультимодальные модели: одна модель позволит обрабатывать весь массив речевых данных и сможет решать максимум задач обработки естественного языка, а генеративный ИИ позволит реализовать то, чего ранее не существовало.

G: Недавно группа ЦРТ объявила о реализации экспериментов с большими языковыми моделями, можете рассказать об этом подробнее?

Д. Д.: Да, мы одними из первых, еще в прошлом году, реализовали эксперименты по внедрению больших языковых моделей в свои продукты: в речевую аналитику, диалоговых ассистентов, чтобы изучить возможности их применения в контактных центрах. Эксперименты продемонстрировали, что с помощью комбинации больших языковых моделей с другими моделями ряд процессов можно упростить, оптимизировать, они могут помочь создавать диалоговых ассистентов, сокращая ручное обучение, и заменить ряд предыдущих алгоритмов.

Например, мы разработали AI-решение для аналитики и управления коммуникациями в контактных центрах. Алгоритмы машинного обучения позволили увеличить скорость обработки и анализа обращений клиентов, а их совмещение с нейросетевой моделью GigaChat позволило нам предоставить бизнесу новый функционал: суммаризацию содержания и тематик диалогов, оценку их результативности и тональности. Теперь все данные агрегируются, расшифровываются и анализируются в онлайн-режиме прямо в момент обращения клиента. Это помогает оператору давать более точные ответы на запросы. Новое решение позволяет управлять клиентским



ЕВГЕНИЙ ПАВЛЕНКО

опытом, увеличивает экономическую эффективность и развивает бизнес. Мы ведем работу по усилению целого спектра решений из нашего продуктового портфеля LLM.

G: В каких отраслях речевые технологии показывают максимальную эффективность?

Д. Д.: К инновациям и технологическим новшествам традиционно открыты финтех и телеком. Так, к примеру, в том числе за счет технологий речевой аналитики ЦРТ удалось улучшить ключевые показатели клиентского опыта в контактном центре «Сбера» для корпоративных клиентов: экономический эффект от внедрения за 2023 год составил 189 млн рублей, а индекс удовлетворенности клиентов, включая показатель удобства пользования по состоянию на 2023 год, достиг 81,62%.

G: Как развивается рынок речевой аналитики?

Д. Д.: По доле выручки от внедрения onprem речевой аналитики (которая устанавливается на серверах клиентов) в крупном бизнесе лидирует ЦРТ, что подтверждают независимые исследования. По нашим прогнозам, российский рынок речевой аналитики в среднем вырастет на 15%. Согласно оценке Markets & Markets, 73% рынка Conversational AI (продуктов на базе речевых технологий.— **G**) относится к облачному сегменту, который в ближайшие годы будет расти со среднегодовой скоростью 22% и к 2026 году удвоится. При этом внедрение речевой аналитики в закрытом контуре предприятия останется актуальным для суперкрупного бизнеса с высокой специализацией.

G: С учетом этих трендов растут ли вложения в обеспечение безопасности облачного ПО?

Д. Д.: Безусловно. Облачное ПО менее ресурсозатратно, проще в обслуживании, не требует дополнительных вложений со стороны пользователя. У авторитетных вендоров облачные технологии защищены от угроз, а ПО соответствует современным требованиям безопасности, защите персональных данных. Так что сегодня можно подходить к разворачиванию ПО в облаке.

G: Реализованы ли кейсы по совмещению речевых технологий и LLM, которых ранее не существовало?

Д. Д.: Одна из отраслей-первопроходцев — промышленность. Мы с нашими партнерами — «Сибуром» и «Сбером» — в числе первых в стране реализовали внедрение LLM в промышленности, договорились о реализации практических кейсов применения нейросетевой модели «Сбера» GigaChat. В «Сибуре» — одной из крупнейших интегрированных нефтегазохимических компаний России — уже созданы AI-помощники. С помощью AI ассистента-диагноста инженер может вести диалог о причине неисправности оборудования. Сотрудник в текстовом виде описывает аномалии в работе оборудования, а обученный AI-помощник дает релевантный ответ или формулирует гипотезы о причинах нехарактерного поведения агрегатов, а также возможные варианты устранения неисправности. И это лишь один из кейсов.

G: С каким запросом к вам обращается бизнес?

Д. Д.: Цель у многих одна: оптимизировать с помощью AI-решений бизнес-процессы, увеличить производительность труда, сократить затраты и при этом предоставлять лучший сервис.

G: Можно ли решить эти вопросы, внедряя технологии (включая LLM) самостоятельно?

Д. Д.: Большие языковые модели — технология с огромным потенциалом, но для того чтобы получить реальные эффекты, важно понимать специфику крупного бизнеса, встраивать решения в производственные процессы, трансформируя их. Необходимо тесное взаимодействие бизнеса и разработчиков модели, чтобы решения были предсказуемыми, безопасными и в полной мере учитывали особенности производства. Лучших эффектов можно добиться, опираясь на опыт технологического партнера.

G: Можете ли детальнее рассказать об этом?

Д. Д.: Одна из целей ЦРТ — демократизация доступа к возможностям LLM для бизнеса разного уровня зрелости. Наши клиенты видят в нас технологического партнера, с помощью которого они могут быстрее и проще получить доступ к LLM, приобрести экстравозможности. Например, команда ЦРТ применила NLU-подходы и представила решение, которое совмещает в себе экспертизу в области обработки речи с возможностями GigaChat, при этом расширила их с помощью RAG (Retrieval-Augmented Generation — генерации, дополненной поисковой выдачей.— **G**). Такая интеграция помогает GigaChat генерировать ответ с использованием информации из внешней базы знаний. На базе нового инструмента можно создать простой интерфейс, который позволит сотрудникам задать вопрос на естественном языке и получить ответ, сформулированный с учетом всех имеющихся в компании информационных систем: CRM, 1С, HR, ГОСТов, внутренних регламентов или ресурсов, включая информацию из Excel и других видов табличных документов.

G: Бизнес, обращаясь к технологическому партнеру, получает доступ не только к технологии, но и к полноте опыта?

Д. Д.: Именно. Наши клиенты ценят, что, обратившись в ЦРТ, они получают доступ не просто к продукту, а более чем к 30-летней экспертизе и пониманию физики корпоративных процессов крупного бизнеса, собственному R&D-центру. Такой комплексный подход позволяет нашим клиентам, с одной стороны, создавать прорывные решения, которых до этого на рынке не существовало, а с другой — опираться на прочный фундамент с пониманием дальнейшего развития AI-решения в зачастую сложной инфраструктуре. Это помогает добиться лучших результатов в конкретном кейсе и развивает рынок в целом. ■

В ПЕРСПЕКТИВЕ ВСЕ БОЛЬШЕ РЕЧЕВЫХ ТЕХНОЛОГИЙ БУДЕТ ОБЪЕДИНЯТЬСЯ В БОЛЬШИЕ МУЛЬТИМОДАЛЬНЫЕ МОДЕЛИ: ОДНА МОДЕЛЬ ПОЗВОЛИТ ОБРАБАТЫВАТЬ ВЕСЬ МАССИВ РЕЧЕВЫХ ДАННЫХ И СМОЖЕТ РЕШАТЬ МАКСИМУМ ЗАДАЧ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА, А ГЕНЕРАТИВНЫЙ ИИ ПОЗВОЛИТ РЕАЛИЗОВАТЬ ТО, ЧЕГО РАНЕЕ НЕ СУЩЕСТВОВАЛО