

Особенности применения искусственного интеллекта для бизнеса в сфере услуг

Заявка № 1674123

За последние 5 лет интеграция искусственного интеллекта в бизнес-процессы стала определяющей характеристикой экономического развития во всем мире. Для Российской Федерации эта технологическая трансформация имеет особое значение в условиях геополитической нестабильности, санкций, подорожания компьютерных компонентов и необходимости развития собственных технологических возможностей.

Сфера услуг включает в себя банковское дело, розничную торговлю, телекоммуникации, консалтинг и так далее. Сфера услуг заслуживает пристального внимания по трем причинам. Во-первых, на сферу услуг приходится около 60% ВВП России, и она обеспечивает занятость наибольшей части населения. Во-вторых, применение ИИ в секторе услуг (чат-боты, поисковые системы, автоматизация процессов) является наиболее распространенным и показательным, можно проследить и сравнить, как эти технологии используются в других странах. В-третьих, B2C (business to customer) услуги, являются наиболее наглядным проявлением влияния ИИ для рядового пользователя, общественное восприятие искусственного интеллекта формируется от взаимодействия с именно этими услугами.

Консалтинг-компания «Яков и партнеры» в своем исследовании «Artificial Intelligence in Russia — 2025: Trends and Outlook» выявили, что более 70% российских компаний уже используют искусственный интеллект в той или иной мере для выполнения своих задач. Основными направлениями являются компьютерное зрение, обработка языка и рекомендательные системы. Большая часть таких систем основана на готовых решениях с открытым кодом, которые дообучают для определенных задач. Экономический эффект от применения ИИ составил в среднем 3,3% по оценке авторов, в то же время ожидаемый эффект оценили на уровне 10%. Таким образом, имплементация ИИ на данном этапе не оправдала ожиданий, но оказали значимый положительный эффект, доказывая необходимость развития данной технологии.

Начальный этап внедрения искусственного интеллекта российские компании смогли пройти достаточно успешно, адаптируя готовые решения с открытым кодом под локальные задачи. Но на фазе масштабирования они могут столкнуться с серьезными препятствиями. Главной опасностью для будущего развития является сложность наращивания вычислительных мощностей. Успешный проект, обслуживающий сотни или даже тысячи запросов ежедневно, требует для своей работы специфические серверные мощности с продвинутыми графическими процессорами и оперативной памятью.

Санкции, введенные против РФ в 2022 году, привели к сокращению импорта этих критически важных компонентов - современных графических процессоров и чипов для серверов. Американская компания NVIDIA, мировой лидер в производстве микропроцессоров для искусственного интеллекта, полностью прекратила поставки своей продукции в Россию. Это означает, что компании, желающие масштабировать свои ИИ системы, не могут просто заказать необходимое оборудование. Единственным каналом остается "серый импорт" через третьи страны, что сильно ограничивает возможности развития.

На рынке доступно лишь некоторое оборудование, в ограниченном количестве и последнего поколения, что усложняет процесс встраивания их в существующую инфраструктуру. Даже без этих ограничений стоимость компьютерных компонентов сильно выросла за последние годы, что делает развитие ИИ систем для среднего и малого бизнеса нерентабельным.

Все это создает критическую зависимость от ограниченных облачных мощностей. Компании вынуждены арендовать вычислительные ресурсы у крупных российских игроков, например Yandex Cloud и Cloud.ru. Хотя это позволяет временно решить проблему дефицита мощностей, в долгосрочной перспективе это может привести к монополизации рынка или "цифровому феодализму". Таким образом, технологические санкции превратили процесс масштабирования ИИ в РФ из инженерной задачи в сложнейшее логистическое и финансовое испытание.

Для решения этих уникальных задач российским компаниям необходимо адаптировать процесс масштабирования. Во-первых, создания универсальных моделей, компаниям необходимо сфокусироваться на узких, тщательно отобранных бизнес-функциях. Такой подход предполагает дообучение моделей с открытым кодом под конкретные задачи, что позволяет добиться приемлемого качества при минимальных требованиях к инфраструктуре

Во-вторых, технологический разворот на Восток. Прямое сотрудничество с китайскими облачными провайдерами может открыть доступ к альтернативным источникам вычислительных мощностей. Хотя это создает новую зависимость от китайской инфраструктуры, в краткосрочной перспективе данный канал поможет получить ресурсы для масштабирования и снизит риски «цифрового феодализма».

В-третьих, необходимо поставить в приоритет эффективность за счет оптимизации. В долгосрочной перспективе, учитывая ограничения, наложенные на российские компании, и низкий уровень инвестиций, только модели с высоким уровнем адаптации могут быть полноценно масштабированы. Способность эффективно работать на разнородном и даже устаревшем оборудовании будет определяющей характеристикой для будущего гиганта ИИ на российском рынке.