

Искусственный интеллект в России – 2025: тренды и перспективы

Максим Болотских, Марина Дорохова, Иван Серов,
Максим Шуркин, Сабина Бабаева, Даниил Ермолаев

Артур Самигуллин, Елена Белоброва, Николай Савушкин,
Сергей Овчаренко, Алексей Колесов

Москва, 2025



Содержание

Введение	4
<hr/>	
Краткое содержание	7
<hr/>	
Текущий статус внедрения ИИ-технологий в России	11
Уровень развития технологий искусственного интеллекта в России	11
Подход к внедрению технологий	20
<hr/>	
Ожидаемый финансовый эффект от внедрения ИИ	26
Распределение эффекта по технологиям	26
Распределение эффекта по отраслям	28
<hr/>	
Генеративный ИИ	32
Развитие технологии генеративного ИИ	32
Ключевые тренды и изменения в технологии генеративного ИИ	36
Основные области внедрения технологии генеративного ИИ	40
Рынок B2B-решений технологии генеративного ИИ	49
Внедрение решений генеративного ИИ в организациях	55
Будущее и перспективы развития генеративного ИИ	60
<hr/>	
Обработка естественного языка (NLP) и речевые технологии	64
Развитие NLP и речевых технологий	64
Ключевые тренды и изменения в NLP и речевых технологиях	66
Основные области внедрения речевых технологий и NLP	67
Внедрение решений речевых технологий и NLP в организациях	71
Рынок B2B-решений речевых технологий и NLP	74
Будущее и перспективы развития речевых технологий и NLP	76

Компьютерное зрение (CV)	78
Развитие технологии CV	78
Ключевые тренды и изменения в технологии CV	80
Основные области внедрения технологии CV	81
Внедрение решений CV в организациях	85
Рынок B2B-решений технологии CV	90
Будущее и перспективы развития технологии компьютерного зрения	92
<hr/>	
Рекомендательные системы (RecSys)	94
Развитие технологии RecSys	94
Ключевые тренды и изменения в технологии RecSys	96
Основные области внедрения технологии RecSys	97
Внедрение решений RecSys в организациях	102
Рынок B2B-решений технологии RecSys	106
Будущее и перспективы развития технологии рекомендательных систем	108
<hr/>	
Заключение	110
<hr/>	
Примечания	117

Введение

Этот отчет — третий в серии ежегодных исследований «Яков и Партнёры» об искусственном интеллекте, и в этом году он проводится совместно с компанией «Яндекс». В 2023 г. фокус был на генеративном ИИ в России, в 2024 г. — на сравнительном анализе экосистем ИИ стран БРИКС+. В 2025 г. объектом анализа выступают четыре ключевые технологии: генеративный ИИ с акцентом на ИИ-агентах, обработка естественного языка (NLP) и речевые технологии, компьютерное зрение (CV), рекомендательные системы (RecSys).

Выбор тематики объясняется разноуровневой зрелостью направлений ИИ. Речь идет не только о генеративных моделях, но и о технологиях, которые уже много лет внедряются и масштабируются, — речевых технологиях и NLP, компьютерном зрении (CV) и рекомендательных системах (RecSys). В то же время сегмент генеративного ИИ в 2024–2025 гг. проходит этап перехода от интереса к практическому внедрению. ИИ все чаще понимается не как набор точечных решений, а как платформенный слой, обеспечивающий работу сервисов с большими массивами данных и разнородными источниками.

В этих условиях анализ текущего состояния рынка особенно уместен: он позволяет одновременно зафиксировать зрелые практики и источники эффекта технологий с длительным циклом развития и оценить траекторию ускоренного роста генеративного ИИ и ИИ-агентов. Поэтому исследование охватывает несколько технологических блоков, а не ограничивается генеративными моделями.

Исходя из этого, в рамках исследования рассмотрены следующие аспекты:

- Всесторонне проанализирован уровень зрелости каждой ИИ-технологии в России — как с позиции разработки, так и с позиции практического внедрения, включая оценку уровня развития рынка и экономических эффектов, которые приносит ИИ бизнесу и экономике в целом.
- Сформирован прогноз развития рынка ИИ-технологий в Российской Федерации до 2030 г.
- Оценена глубина использования четырех ключевых технологий в различных отраслях и функциональных областях (маркетинг и продажи, клиентский сервис, операционная эффективность, риск-менеджмент и др.).
- Охарактеризована структура и динамика рынка вендоров ИИ — их фокус, уровень зрелости, модели сотрудничества, используемая инфраструктура и экономика решений.

- Проанализировано отношение пользователей к технологиям ИИ и уровень их доверия к инновациям.
- Определены технологические и управленческие приоритеты на горизонте до 2030 г.

Для комплексной оценки текущего состояния развития искусственного интеллекта в России «Яков и Партнёры» совместно с «Яндексом» в 2025 г. провели три масштабных исследования.

1. Первое — опрос технических директоров (Chief Technology Officers, СТО) 150 крупнейших компаний России, представляющих 16 отраслей экономики: нефть и газ, FMCG, розничную торговлю и e-commerce, банкинг и страхование, металлы и горную добычу, машиностроение, транспорт и логистику, автомобильную промышленность, электроэнергетику, сельское хозяйство, химию и нефтехимию, телеком и медиа, ИТ и технологии, строительство и недвижимость, а также медицину и здравоохранение. Опрос позволил оценить уровень внедрения технологий ИИ по функциям бизнеса, источники эффекта, барьеры для реализации, а также понять восприятие ИИ-технологий со стороны бизнеса, готовность инвестировать и их значимость для развития.
2. Второе — исследование рынка поставщиков ИИ-решений: опрошены топ-150 российских вендоров, работающих в сегментах генеративного ИИ, компьютерного зрения (CV), рекомендательных систем (RecSys), а также NLP и речевых технологий. Фокус — предложения и специализация, модели внедрения, профиль поставщиков (количество лет на рынке, выручка, штатная численность), ключевые индустрии, в которых происходит продажа решений, а также их дальнейшая перспектива.
3. Третье — опрос пользователей с участием свыше 3,5 тыс. жителей России, целью которого было оценить уровень осведомленности и вовлеченности пользователей в тематику ИИ в целом и рассматриваемых технологий в частности: доверие, готовность к использованию, оценка рисков и пользы.

Дополнительно проведена серия глубинных интервью с руководителями и экспертами компаний, формирующими повестку в сфере ИИ в России, что позволило дополнить количественные данные качественными наблюдениями и практическими кейсами.

Комбинация трех опросов и глубинных интервью позволяет сопоставить перспективу корпоративных ИТ-лидеров, поставщиков решений и конечных пользователей, выявить точки роста и ограничения, а также зафиксировать лучшие практики использования конкретных кейсов на базе ИИ.

Полученные результаты и аналитические выводы предназначены для широкого круга читателей, вовлеченных в развитие и применение технологий искусственного интеллекта:

- Руководителям — как основа для постановки приоритетов по внедрению ИИ и управлению рисками.
- Технологическим и продуктовым командам — как ориентир по архитектуре сервисного слоя, метрикам качества и эффекту.
- Вендорам и партнерам — для соотнесения портфеля решений с отраслевым спросом.
- Регуляторам и академическому сообществу — для диалога о стандартах, данных, инфраструктуре и кадрах.
- Широкому кругу заинтересованных специалистов и читателей — как возможность системно погрузиться в тематику ИИ, понять тенденции, масштабы внедрения и влияние технологии на экономику и общество.

Исследование имеет следующую структуру по разделам.

В разделе «Текущий статус внедрения ИИ-технологий в России» читатель найдет обзор зрелости разработки и практического применения ИИ, отраслевые различия и динамику внедрения, модели инвестиций и источники эффекта, а также организационные подходы к внедрению (ответственность, роли, форматы сотрудничества с вендорами, этапы проектов).

Раздел «Ожидаемый финансовый эффект от внедрения ИИ» описывает методику количественной оценки вклада ИИ, распределение эффекта по технологиям и отраслям, сравнение абсолютного и удельного влияния на EBITDA и горизонты ожидаемых изменений до 2030 г.

Далее следуют четыре технологических обзора: генеративный ИИ, обработка естественного языка (далее — NLP) и речевые технологии, компьютерное зрение (далее — CV) и рекомендательные системы (далее — RecSys). Каждый обзор построен по единой логике: этапы развития технологии; ключевые тренды и изменения; основные области внедрения; практики внедрения в организациях; рынок B2B-решений; будущее и перспективы развития. Такой формат позволяет быстро сопоставить технологии между собой и увидеть прикладные сценарии использования.

В «Заключении» представлена сводная рамка по каждой технологии, роль ИИ в развитии РФ, а также практические рекомендации для государства, компаний и людей.

Мы рассчитываем, что представленные материалы будут полезны как практический ориентир для принятия решений и планирования. Надеемся, что результаты исследования помогут ускорить переход от экспериментальных инициатив к устойчивому, безопасному и экономически обоснованному применению ИИ.

Краткое содержание

В отчете «Искусственный интеллект в России — 2025: тренды и перспективы», подготовленном «Яков и Партнёры» совместно с компанией «Яндекс», выводы опираются на три крупных массива данных: опрос 150 СТО из 16 отраслей, опрос 150 российских вендоров ИИ и пользовательский опрос свыше 3,5 тыс. респондентов, которые дополнены глубинными интервью с отраслевыми лидерами. Такой дизайн позволяет сопоставить взгляды корпоративных ИТ-руководителей, поставщиков ИИ-решений и конечных пользователей.

Текущий статус внедрения ИИ в России характеризуется сильной инженерной базой и разрывом между уровнем разработки и глубиной интеграции в бизнес-процессы. Все крупные компании уже прошли стадию пилотов хотя бы по одной из четырех технологий, а в среднем по каждой технологии пилоты идут примерно у 88% компаний. Продвинутые в ИИ отрасли — ИТ и технологии, телеком и медиа, e-commerce, банкинг и страхование, где ускорению способствуют развитая инфраструктура, высокая конкуренция и наличие внутренних компетенций. В целом активная фаза интеграции современных направлений ИИ (генеративный ИИ, RecSys, CV, NLP и речевые технологии) началась в последние два года, что указывает на переход от интереса к формированию устойчивых практик внедрения. Ключевой запрос пользователей к ИИ — прикладная польза: модели должны решать повседневные задачи, поэтому встраивание технологий в привычные сервисы и интерфейсы — необходимое условие для дальнейшего роста рынка.

Ожидаемый экономический эффект от ИИ к 2030 г. оценивается в 7,9–12,8 трлн руб. в год, что соответствует до 5,5% прогнозного ВВП. Эффект складывается не только из оптимизации затрат, но и из новой выручки за счет продуктовых инноваций и трансформации бизнес-моделей. В технологическом разрезе наибольший вклад формируют генеративный ИИ и компьютерное зрение, в отраслевом — наибольшая удельная эффективность фиксируется в digital-ориентированных индустриях: e-commerce, телеком и медиа, ИТ и технологии. Компании, выстроившие процесс измерения эффекта, видят основной вклад не как сокращение ФОТ, а как рост выручки и маржинальности за счет персонализации, улучшения качества сервиса и ускорения вывода продуктов.

Источник: открытые источники,
анализ «Яков и Партнёры»

Ожидаемый экономический эффект
от ИИ к 2030 г. оценивается в

7,9–12,8
трлн руб. в год

Генеративный ИИ за два года перешел от экспериментальных запусков к массовому использованию: доля компаний, применяющих технологию хотя бы в одной функции, достигла 71% в 2025 г. (на 17 п. п. выше, чем годом ранее). Наибольший масштаб внедрения — во внутренних коммуникациях, в маркетинге и продажах и в клиентском сервисе; бизнес отмечает вклад генеративного ИИ в создание новых продуктов и закладывает дальнейший рост инвестиций, делая технологию приоритетной в ИТ-бюджетах. В России, помимо проприетарных моделей, активно пользуются решениями на базе open-source-моделей: они помогают снижать издержки и ускорять запуск сервисов и фичей для пользователей и бизнеса; без этого пользовательский интерес мог бы сместиться в пользу иностранных ИИ-сервисов. Отдельно подчеркивается, что в стране есть сопоставимые по качеству конкуренты крупнейшим LLM мира (включая DeepSeek и ChatGPT) — Alice AI (предыдущее поколение нейросети называлось YandexGPT), GigaChat, что подтверждает зрелость экосистемы и уровень локальной экспертизы. Одновременно развивается класс ИИ-агентов, расширяющий сценарии от простых вопросов и ответов к многошаговой автономной деятельности и интеграции с корпоративными системами.

NLP и речевые технологии демонстрируют высокую степень зрелости и являются наиболее узнаваемыми среди пользователей (53% пользователей знакомы с технологией против 39% в среднем по другим технологиям); основная область применения — клиентский сервис (68% компаний применяют технологию в этой функции), где решения уже стали базовой инфраструктурой обработки обращений и диалогового взаимодействия, а также производство и маркетинг. Компании в продвинутых в ИИ отраслях ожидают ощутимого вклада NLP на горизонте года в EBITDA (3,6%) и планируют поддерживать стабильные инвестиции, усиливая связку с генеративным ИИ там, где требуется более гибкое понимание контекста и генерация сложных ответов.

Компьютерное зрение остается одним из ключевых драйверов экономического эффекта (17% совокупного эффекта) благодаря автоматизации визуального контроля, безопасности и операций в физических процессах, прежде всего в производстве (67% компаний применяют технологию в этой функции). Внедрение CV в производственные контуры, контроль качества и предиктивное обслуживание оборудования формирует устойчивый вклад в операционную эффективность и будет продолжать масштабироваться.

Рекомендательные системы — одно из наиболее зрелых направлений (54% компаний внедрили технологию более 5 лет назад против 30% в среднем по другим отраслям), имеющее самую длительную историю внедрения и отложенную экономику. Порог входа и капитальные затраты ниже, а путь к измеримому результату короче, чем у других ИИ-технологий. Их ключевая ценность — усиление работы маркетинга и продаж и клиентского сервиса (77 и 69% компаний применяют технологию в этих функциях соответственно). В e-commerce доля инвестиций выше среднего (4,0% против 1,7%), а эффект по компаниям достигает около 1% EBITDA. В ближайшие годы ожидаются эксплуатационная донастройка и интеграция с генеративным ИИ для динамической генерации контента, при этом классические алгоритмы сохранят роль базового слоя прогнозной аналитики.

Зачастую технологии ИИ делят на четыре независимые области — генеративный ИИ, NLP и речевые технологии, CV и RecSys. Проведенный анализ ИИ-технологий показывает, что эту схему следует пересмотреть: генеративный ИИ — не четвертая вертикаль, а горизонтальная платформа, которая интегрируется в оставшиеся три области изнутри. Современные модели становятся активными исполнителями: они не только отвечают, но и извлекают и проверяют информацию, обращаясь к внешним базам и API, инициируя набор действий. На этом фоне быстро развивается класс ИИ-агентов и их координация между собой. Происходит переход от множества узкоспециализированных моделей к foundation-моделям общего назначения: если раньше для каждой задачи требовалась своя модель (диалог, анализ изображений, рекомендации), то теперь одна базовая модель, дополненная инструментами и данными, покрывает широкий набор сценариев использования. Именно это и определяет ключевые структурные сдвиги ближайших лет.

В совокупности данные показывают: российская ИИ-экосистема вступила в фазу масштабирования, где зрелые технологии — CV, RecSys, NLP и речевые технологии — обеспечивают предсказуемый эффект, а генеративный ИИ и агентские системы становятся ускорителем новых продуктов и бизнес-моделей. В масштабах страны достижение эффекта от ИИ возможно лишь при плотном взаимодействии образовательных организаций, государства и частных компаний: образовательная система формирует у людей практические навыки и культуру осознанного взаимодействия с моделями, бизнес выступает поставщиком и интегратором этих моделей, а государство обеспечивает достаточную и комфортную инфраструктуру и регуляторные условия для их полномасштабного развертывания.

Текущий статус внедрения ИИ-технологий в России

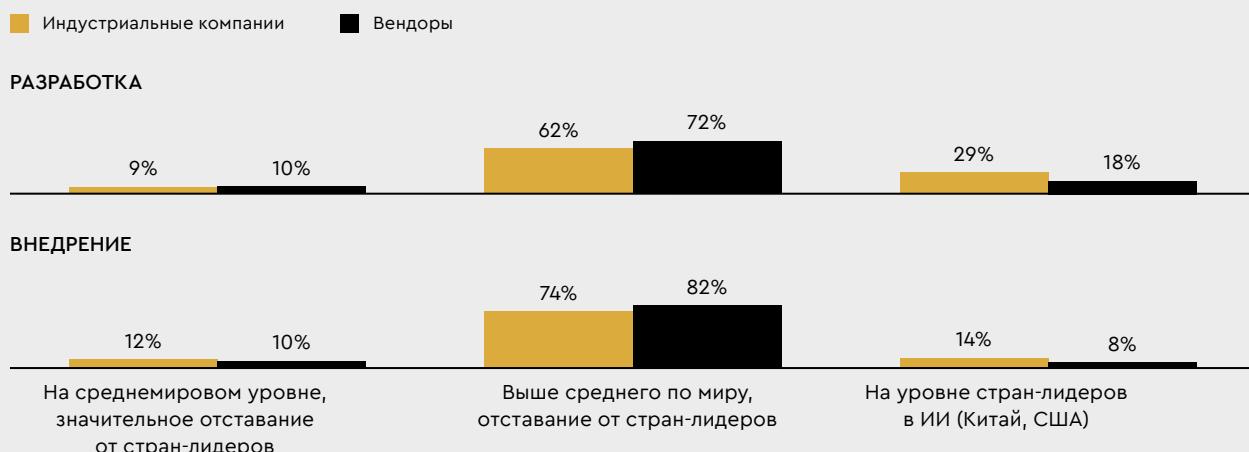
Уровень развития технологий искусственного интеллекта в России

Согласно результатам опросов СТО и вендоров, в среднем по всем технологиям около 90% респондентов — как среди технологических вендоров, так и среди индустриальных компаний — оценивают уровень развития и внедрения искусственного интеллекта в России как «выше среднего по миру» или сопоставимый с уровнем стран-лидеров, таких как США и Китай.

При этом оценки по направлениям разработки и внедрения демонстрируют определенный разрыв: около 18% вендоров и 29% индустриальных компаний считают, что Россия уже достигла уровня мировых лидеров по разработке ИИ-решений, тогда как по внедрению такие оценки дают 8 и 14% респондентов соответственно.

Это подчеркивает разрыв между развитием инженерных компетенций и реальной интеграцией решений в бизнес-процессы. Несмотря на высокие оценки уровня разработки технологий, фактическое внедрение все еще отстает от мировых лидеров, что свидетельствует о потребности в ускорении перехода от разработки к массовому применению решений ИИ.

Восприятие уровня разработки и внедрения ИИ в России, % компаний и вендоров



Источник: опрос индустриальных компаний (блок 1. Контекст, вопрос 1.1. Как вы оцениваете текущий уровень разработки и внедрения технологий ИИ в России?); опрос вендоров (блок 1. Контекст, вопрос 1.1. Как вы оцениваете текущий уровень разработки и внедрения технологий ИИ в России?)

Отдельно стоит отметить, что в опросе среди населения России половина опрошенных оценивают позиции страны в области разработки и внедрения ИИ как «выше среднемирового». Однако среди специалистов ИТ-отрасли эта доля значительно выше и достигает 69%, что подчеркивает более высокую оценку ИТ-сообществом достигнутого прогресса.

Период внедрения технологий искусственного интеллекта

Несмотря на высокий уровень интереса к технологиям искусственного интеллекта, активное внедрение современных ИИ-направлений — генеративных моделей, рекомендательных систем, компьютерного зрения и NLP-решений — в российских компаниях началось сравнительно недавно. В последние два года внедрение таких решений ускорилось, и около 70% компаний сфокусировались на интеграции этих технологий в свои бизнес-процессы. Этот сдвиг стал возможен благодаря появлению генеративного ИИ, который значительно сократил время на внедрение и начал оказывать более заметный эффект. Это указывает на то, что рынок находится на стадии формирования зрелых практик применения ИИ, когда эффект от интеграции технологий только начинает масштабироваться.

Самая молодая технология — генеративный ИИ, его внедрение началось около года назад; самая зрелая — рекомендательные системы, активно применяемые уже почти три года

Среди четырех направлений самой «молодой» остается технология генеративного искусственного интеллекта — в среднем около 1 года с момента первых запусков. За ней следуют речевые технологии и NLP (1,7 года) и компьютерное зрение (2,5 года). Наибольший срок насчитывают рекомендательные системы — в среднем почти 3 года с момента их первых внедрений, что делает их наиболее зрелой ИИ-технологией на российском рынке, согласно ответам респондентов. Отрасли со старшим профилем (автопром, энергетика, медицина) раньше остальных начали с прикладных задач (компьютерное зрение/речевые технологии на производственной линии и в клинической документации), поэтому средний возраст технологий у них выше. Напротив, в e-commerce и банкинге новая волна — генеративный ИИ и разговорные интерфейсы на фронте — стартовала лишь в 2023–2024 гг., что омолаживает метрики.

Во внедрении ИИ-технологий прослеживается единый паттерн: лидеры в раннем внедрении одной технологии, такие как ИТ и телеком, как правило, оказываются первыми и в других направлениях. Так, в этих секторах внедрение рекомендательных систем началось в среднем 3–3,5 года назад, а решений на базе компьютерного зрения — около 2,5–3 лет назад. По генеративному ИИ эти отрасли также остаются впереди: срок активного использования технологии составляет в среднем 1,5–2 года, тогда как в менее инновационных индустриях, таких как FMCG и металлы и горная добыча, — менее года. В транспорте и логистике ранний старт компьютерного зрения дополнительно объясняется приоритетом безопасности и непрерывности операций.

Более высокая скорость внедрения в ИТ и телеком объясняется некоторыми факторами: развитой цифровой инфраструктурой, наличием внутренних компетенций в области данных и алгоритмов, а также высокой конкуренцией за клиента, стимулирующей компании оперативно тестировать и масштабировать новые технологические решения.

В то же время развитие ИИ в отрасли не всегда определяется исключительно ее инновационностью. Так, банкинг и страхование имеют более поздний старт внедрения, что обусловлено высокой степенью регулирования и необходимостью обеспечения безопасности при работе с чувствительными данными. Однако, несмотря на это, банки за короткий срок стали одной из самых продвинутых отраслей в области ИИ благодаря высокой конкуренции на рынке и развитию финтеха, которые стимулировали внедрение инновационных технологий и ускорили процесс цифровизации. Дополнительно часть практик в продвинутых в ИИ отраслях респонденты могли не отнести к рекомендательным системам (например, CRM/таргетинг), статистически «омолодив» видимый старт технологии.

Прошедшее время после внедрения первой технологии, годы

■ Более 3 лет ■ От 2 до 3 лет ■ От 1 до 2 лет ■ Менее года

	Генеративный ИИ	Речевые технологии и NLP	CV	RecSys	Среднее кол-во функций
ИТ и технологии	1,58	3,13	3,37	3,87	3,0
Электроэнергетика	1,50	2,50	3,17	2,93	2,5
Автомобильная промышленность	1,50	2,08	3,17	3,17	2,5
Транспорт и логистика	1,25	2,11	2,46	3,72	2,4
Медицина и здравоохранение	1,13	2,21	2,98	3,17	2,4
Телеком и медиа	1,35	1,63	2,75	3,50	2,3
Банкинг и страхование	0,99	2,16	2,29	2,87	2,1
E-commerce	0,91	1,50	2,00	3,29	1,9
Строительство и недвижимость	1,31	1,69	1,93	2,57	1,9
Сельское хозяйство	0,75	1,13	3,17	2,35	1,8
Химия и нефтехимия	1,50	1,50	2,33	1,92	1,8
Ритейл	0,95	1,23	1,88	3,02	1,8
Машиностроение	0,63	1,13	2,35	2,13	1,6
Нефть и газ	0,95	1,57	1,94	1,75	1,6
Металлы и горная добыча	0,76	1,07	1,97	2,25	1,5
FMCG	0,47	1,13	1,83	2,50	1,5
Среднее по технологиям	1,1	1,7	2,5	2,8	2,0

Источник: опрос индустриальных компаний (блок 1. Контекст, вопрос 1.3. Какой период прошел с момента первого внедрения технологии в бизнес-процессы?)

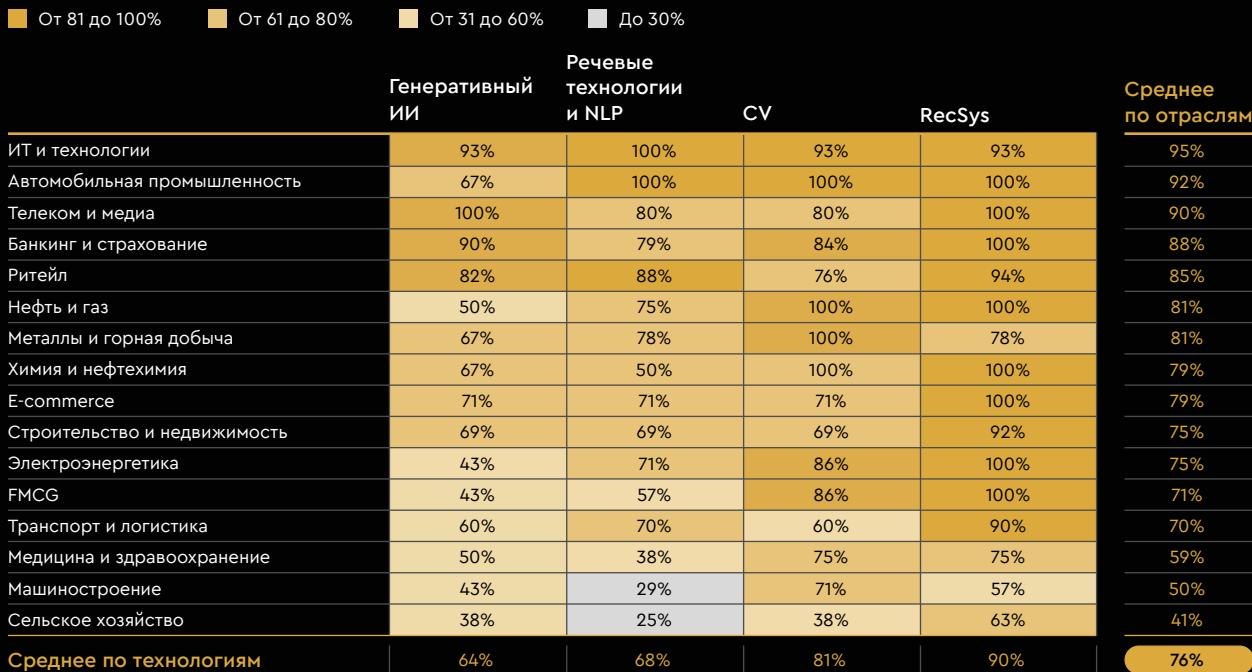
Отраслевая зрелость внедрения технологий искусственного интеллекта

Все крупные компании России уже провели пилоты хотя бы по одной из ИИ-технологий

Несмотря на то что массовое внедрение технологий искусственного интеллекта в России началось относительно недавно, их использование уже охватывает широкий круг отраслей и компаний. Внедрение всех четырех технологий — генеративного ИИ, речевых технологий и NLP, компьютерного зрения и рекомендательных систем — происходит повсеместно. По результатам опроса СТО 150 крупнейших компаний России, по каждой из четырех технологий ИИ в среднем 88% компаний запустили пилотные проекты. При этом каждая из компаний провела пилот хотя бы по одной из технологий, что указывает на переход рынка от стадии интереса к практическому применению и тестированию конкретных сценариев использования ИИ.

Почти все отрасли демонстрируют высокий уровень внедрения ИИ-решений хотя бы в одной из функций внутри организации: средний показатель по четырем ключевым направлениям составляет около 80%. Исключение составляют сельское хозяйство, где уровень внедрения достигает лишь 40%, медицина и здравоохранение (59%) и машиностроение (около 50%), что объясняется более консервативным подходом к инновациям в этих отраслях. В то время как ИТ, телеком и банковский сектор, напротив, характеризуются ускоренной разработкой и интеграцией новых технологий, что позволяет им достигать максимальных показателей — от 90 до 95% компаний уже используют ИИ в своей деятельности.

Компании РФ, которые внедрили ИИ-технологии хотя бы в одну функцию, % компаний



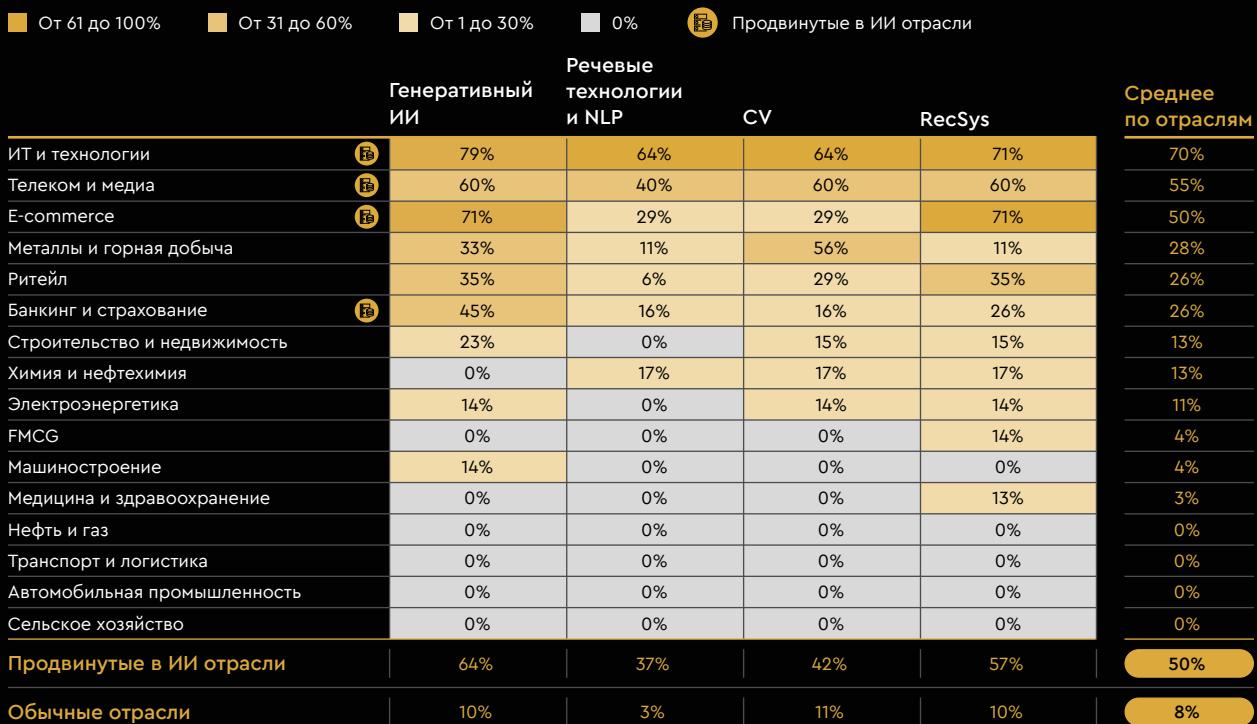
Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.1. В каких функциях внедряется данная технология и в какой степени?)

Продвинутые в развитии искусственного интеллекта отрасли

Продвинутые в ИИ отрасли — ИТ и технологии, телеком и медиа, e-commerce, банкинг и страхование

Наиболее продвинутыми в использовании ИИ мы считаем ИТ и технологии, телеком и медиа, e-commerce, банкинг и страхование. Эти сектора уже перешли от точечных пилотов к системному применению: масштабируют генеративный ИИ и классические технологии во фронте и бэке, опираются на сильные внутренние компетенции и зрелую инфраструктуру, а высокая конкуренция подталкивает их быстрее превращать эксперименты в эффект на бизнес. Наша оценка подтверждается тем, что именно эти отрасли стабильно занимают верхние позиции в наших сравнениях: они чаще других внедряют ИИ сразу в нескольких функциях, направляют на ИИ большую долю ИТ-бюджета и демонстрируют наивысший зафиксированный эффект. Так, e-commerce находится в числе лидеров по инвестициям и фактическому влиянию на показатели, телеком и медиа — среди первых по масштабу внедрения и темпам развития, банкинг и страхование — в верхней группе по измеряемому эффекту и скорости масштабирования, а ИТ и технологии — в лидерах по охвату функций и глубине экспертизы. Дополнительно этот вывод резонирует с восприятием потребителей, которые именно в этих отраслях чаще замечают активное применение ИИ. Совокупность верхних позиций по ключевым срезам позволяет уверенно квалифицировать перечисленные отрасли как наиболее продвинутые.

Компании РФ, которые внедрили ИИ-технологии в несколько функций, % от общего числа компаний отрасли



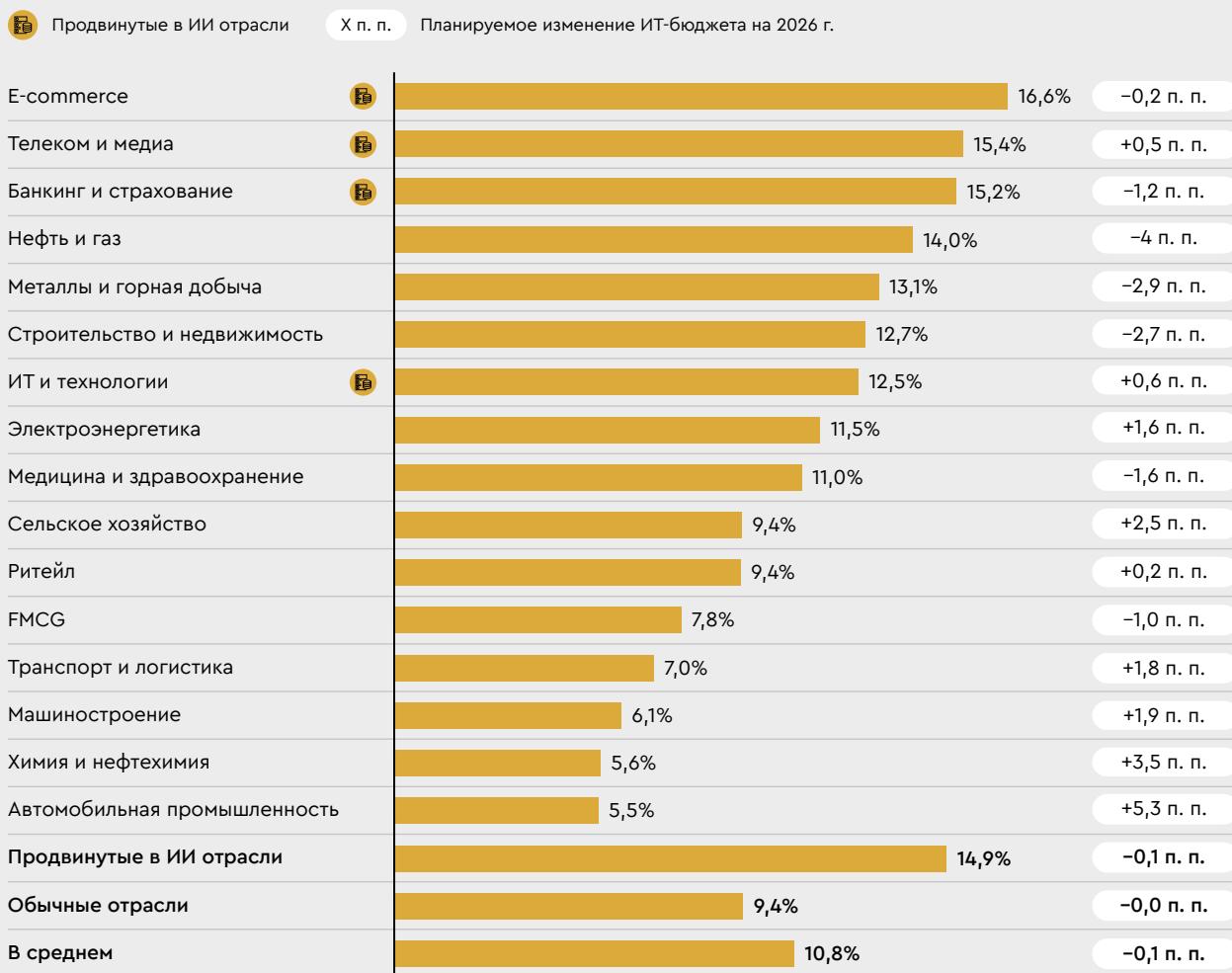
Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.1. В каких функциях внедряется данная технология и в какой степени?)

Инвестиции компаний в ИИ

Продвинутость отраслей в ИИ отражается в трех ключевых факторах — уровне инвестиций в технологии, эффекте от их внедрения и источниках этого эффекта

По результатам опроса СТО, продвинутые в ИИ отрасли направляют на развитие искусственного интеллекта от 13 до 17% годового ИТ-бюджета за последний год, что заметно выше среднего по рынку уровня в 11% и существенно превосходит показатели прочих отраслей, таких как машиностроение, автопром, FMCG и другие, где инвестиции в ИИ ограничиваются 6–9%. Аналогичные различия наблюдаются и в инвестиционных планах: компании продвинутых отраслей закладывают 13–16% ИТ-бюджета на следующий год на развитие искусственного интеллекта, в то время как в прочих отраслях этот показатель не превышает 7–10%. Это подтверждает, что для технологически зрелых секторов ИИ является не отдельным направлением цифровизации, а стратегическим приоритетом развития бизнеса.

Средние инвестиции компаний в технологии ИИ за последний год, в % от годового ИТ-бюджета



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.3. Какой объем средств в % от годового ИТ-бюджета компания уже инвестировала в технологию за последний год? Вопрос 4.5. Какой объем средств в % от ИТ-бюджета компания планирует инвестировать во внедрение технологии на горизонте года?)

Эффект от внедрения ИИ

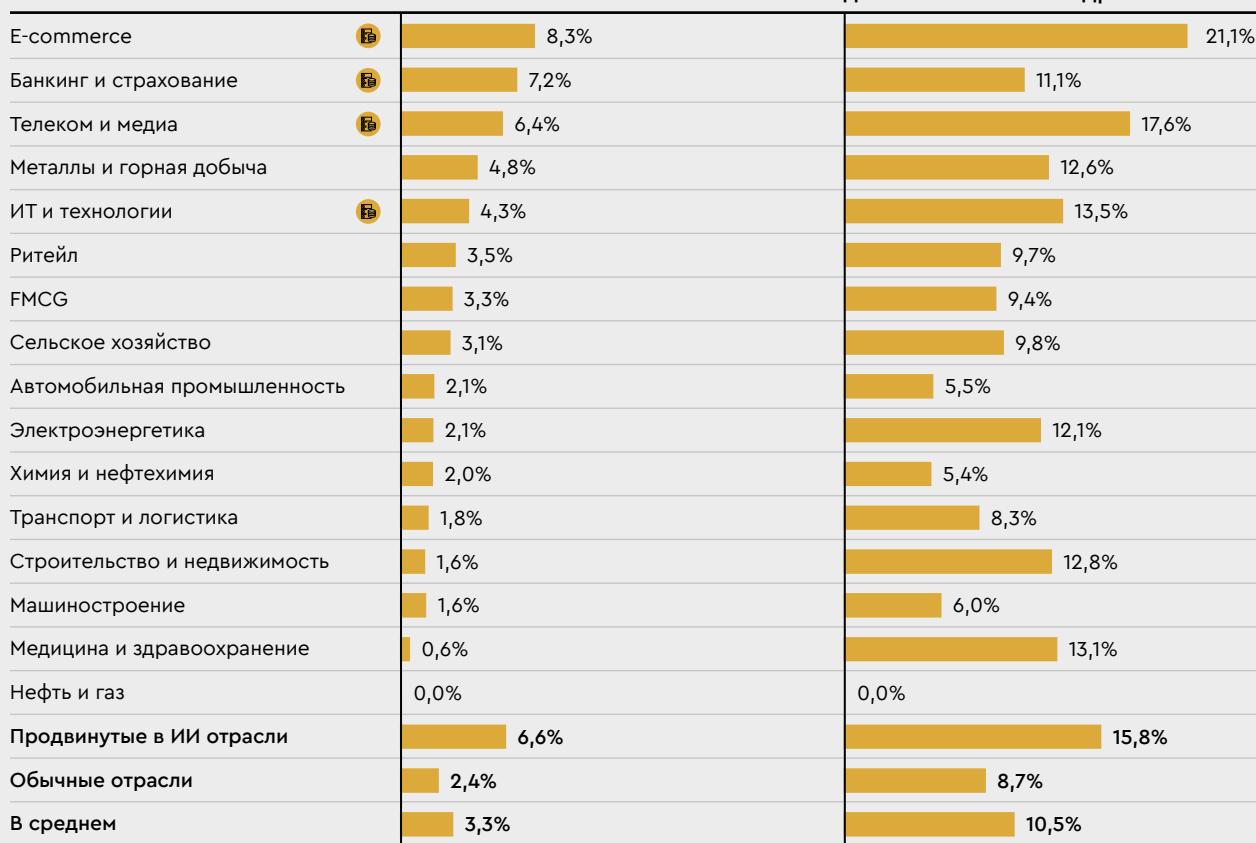
Продвинутые в ИИ отрасли отличаются более высокими возможностями для измерения эффекта от внедрения технологий. Среди этих секторов доля компаний, способных количественно зафиксировать результаты внедрения, значительно выше среднего по рынку. Например, в e-commerce и ИТ до 85% респондентов сообщили о возможности расчета эффекта. При этом эти отрасли также демонстрируют наибольший полученный эффект: влияние искусственного интеллекта на финансовые показатели в этих секторах оценивается в размере до 8% EBITDA. Более того, именно эти отрасли ожидают и максимальный будущий финансовый результат. Совокупный эффект от внедрения ИИ на горизонте года оценивается ими в 13–21% EBITDA, что значительно выше среднего уровня по рынку (11%) и в два-три раза выше, чем в прочих отраслях, где ожидания ограничиваются диапазоном 5–10%. Для продвинутых отраслей ИИ уже стал ключевым драйвером роста и повышения эффективности, тогда как для остальных он пока остается инструментом точечного улучшения отдельных бизнес-процессов.

Эффект от внедрения ИИ в отраслях, % EBITDA

💡 Продвинутые в ИИ отрасли

Средний эффект, полученный
компаниями от внедрения
ИИ за 2024 г.

Средний ожидаемый эффект
для компаний от внедрения ИИ



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.2. Какой финансовый эффект компания уже получила с момента внедрения технологии в % от годовой EBITDA за последний год? Вопрос 4.4. Какой финансовый эффект компания ожидает получить на горизонте года от внедрения технологии в % от EBITDA?)

В сравнении с 2023 г. картина стала более четкой: доля компаний в совокупности, не получающих финансового эффекта от ИИ, снизилась с 32 до 22%, что свидетельствует о повышении зрелости внедрения технологий. Показатель компаний с эффектом на уровне до 5% остался практически неизменным — 68% годом ранее и 69% в последнем периоде. Зато появился ощутимый рост компаний с заметным результатом: доля тех, кто фиксирует влияние ИИ на уровне свыше 5% EBITDA, увеличилась с единичных случаев до 9%.

Не менее важное различие между продвинутыми в ИИ и прочими отраслями наблюдается в источниках финансового эффекта.

Компании из продвинутых секторов — ИТ и технологии, телеком и медиа, e-commerce, банкинг и страхование — значительно чаще связывают влияние искусственного интеллекта с ростом выручки за счет создания новых продуктов и сервисов: этот фактор отмечают 72% компаний продвинутых отраслей против лишь 25% среди прочих. Это отражает уникальный подход к использованию технологий, где создание новых продуктов и сервисов рассматривается как важный источник роста и возможности для дальнейшего развития.

При этом наиболее распространенным источником финансового эффекта от внедрения ИИ для компаний остается сокращение операционных и прочих видов затрат — этот драйвер отмечают почти 90% всех респондентов. Это указывает на практический фокус большинства компаний на повышении эффективности внутренних процессов и управлении издержками.

Среди индустрий именно B2C-секторы — ритейл, e-commerce, телеком и медиа, банкинг и страхование, транспорт и недвижимость — видят в искусственном интеллекте источник роста выручки за счет улучшения работы с клиентской базой. Почти 100% компаний в этих сегментах выделяют этот фактор как ключевой, что обусловлено высокой конкуренцией и необходимостью постоянного совершенствования клиентского опыта и удержания потребителей.

Несмотря на потенциал автоматизации, сокращение затрат на персонал не рассматривается бизнесом как приоритетный источник эффекта. Лишь примерно треть компаний (34% в продвинутых отраслях и 32% в прочих) ожидают снижения фонда оплаты труда за счет внедрения ИИ. Это подтверждает, что компании видят в технологии прежде всего инструмент перераспределения задач и повышения производительности, а не средство сокращения штата.

Кроме того, как показывает практика, компании не готовы снижать планы по найму или оптимизировать число сотрудников для получения реального финансового эффекта на ФОТ. Вместо этого высвободившееся время сотрудников используют для решения новых задач, способствующих росту бизнеса. Мы ожидаем, что перераспределение задач и автоматизация работы, ранее выполняемой человеком, будут в первую очередь происходить в сферах, связанных с работой с текстами (например, юриспруденция, программирование и телемаркетинг). Эксперты подтверждают, что такие изменения уже в ближайшие годы станут основными драйверами изменений в этих отраслях. А к 2030 г. этот процесс достигнет зрелости, охватывая все больше профессий, где ИИ будет способствовать эффективному перераспределению задач и повышению производительности.

Источник эффекта от внедрения ИИ-технологий, % компаний

■ От 81 до 100% ■ От 61 до 80% ■ От 31 до 60% ■ До 30% ■ Продвинутые в ИИ отрасли

		Сокращение прочих видов затрат	Рост выручки за счет улучшения ценности продукта	Рост выручки за счет улучшения работы с клиентской базой	Рост выручки за счет создания новых продуктов	Сокращение затрат на ФОТ
ИТ и технологии	■	100%	50%			75%
E-commerce	■	57%	100%	71%	57%	14%
Строительство и недвижимость		65%	71%	100%	24%	12%
Медицина и здравоохранение		100%	100%	100%	86%	14%
Банкинг и страхование	■	90%	80%	95%	45%	30%
Электроэнергетика		100%	56%	11%	33%	22%
Сельское хозяйство		100%	29%			43%
Металлы и горная добыча		100%	90%	90%		20%
Телеком и медиа	■	100%	67%	67%	67%	33%
FMCG		100%	57%	43%	14%	57%
Транспорт и логистика		100%	71%	57%	29%	57%
Автомобильная промышленность		100%	100%	17%	17%	17%
Ритейл		90%	100%	100%	80%	40%
Химия и нефтехимия		71%	100%	86%	79%	50%
Машиностроение		85%	100%	100%	31%	31%
Нефть и газ		63%	75%	50%	25%	
Продвинутые в ИИ		95%	88%	95%	72%	34%
Обычные отрасли		72%	89%	50%	25%	32%
В среднем		78%	89%	62%	37%	32%

Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.1. В чем компания видит источник финансового эффекта от внедрения технологий?)

Подход к внедрению технологий

Ответственность за внедрение искусственного интеллекта в компании

За внедрение ИИ-решений в 90% случаев отвечают руководители уровня CEO-1

В каждой четвертой компании уже сформирована выделенная ИИ-функция, отвечающая за внедрение решений на базе искусственного интеллекта

Сегодня внедрение искусственного интеллекта рассматривается как стратегическая задача компании, находящаяся в зоне внимания первого уровня руководства. По результатам опроса СТО, за развитие ИИ-инициатив в 90% случаев отвечают руководители уровня CEO-1, что подчеркивает значимость технологии для корпоративной стратегии. При этом около половины компаний (47%) вовлекают в процесс и руководителей уровня CEO-2, обеспечивая функциональное закрепление ответственности и контроль исполнения на уровне подразделений.

Ключевым инициатором и ответственным за внедрение решений на базе искусственного интеллекта в компаниях является ИТ-блок — его указали 89% респондентов. При этом в каждой четвертой компании уже сформирована выделенная ИИ-функция, отвечающая за внедрение и сопровождение решений на базе искусственного интеллекта. Это свидетельствует о переходе от проектного к постоянному организационному формату, где управление ИИ становится частью устойчивой операционной модели компании.

В продвинутых в ИИ отраслях — телеком и медиа, e-commerce, банкинг и страхование — внедрение технологий организовано комплексно: в 75% компаний существуют отдельные департаменты, курирующие развитие ИИ, а в 49% — выделенные специалисты внутри бизнес-функций. Такой подход позволяет интегрировать технологии во все ключевые процессы и масштабировать решения внутри организации. В прочих отраслях модель проще: у 93% компаний существует только один централизованный отдел, ответственный за внедрение ИИ, при этом практически нет выделенных специалистов по внедрению ИИ в бизнес-подразделениях — только 11% компаний имеют таких специалистов. У ИТ-компаний ситуация отличается: отдельные департаменты есть лишь у 45% компаний, поскольку экспертиза уже встроена в основные функции бизнеса.

Модели внедрения ИИ-технологий

По результатам опроса СТО, наиболее распространенной моделью внедрения является использование готовых решений от внешних партнеров — этот формат применяют 78% компаний. При этом около половины всех респондентов отмечают, что адаптируют такие решения совместно с подрядчиком под собственные бизнес-процессы. Наименее распространенной остается полностью внутренняя разработка ИИ-решений — ее реализуют лишь 18% компаний, что отражает высокие требования к ресурсам, компетенциям и инфраструктуре для самостоятельного создания технологий.

Если рассматривать модели внедрения с точки зрения инфраструктуры и развертывания решений, большинство компаний постепенно переходят на облачные или гибридные форматы поставки: 40% респондентов сообщили, что используют ИИ-технологии преимущественно через облачные сервисы, еще 29% предпочитают гибридный формат. Выбор инфраструктурной модели во многом определяется отраслевыми особенностями и требованиями к безопасности данных. Так, в банкинге и страховании около 90% компаний работают исключительно в on-premise-среде из-за регуляторных ограничений и требований по конфиденциальности информации. В то же время в e-commerce ситуация обратная — 71% компаний отметили использование облачных решений. Для этой отрасли ключевым фактором является масштабируемость и гибкость инфраструктуры, что делает облачный подход предпочтительным для быстрого роста и тестирования новых ИИ-сценариев.

Эффективное внедрение ИИ во многом также определяется качеством партнерской экосистемы. Компании все чаще выстраивают сотрудничество как с крупными технологическими корпорациями, так и со специализированными ИИ-вендорами: с технологическими корпорациями работают 66% компаний, с вендорами — 91%. Одновременное сотрудничество с технологическими корпорациями и вендорами позволяет совмещать отраслевую экспертизу и гибкость специализированных разработчиков с инфраструктурными возможностями и масштабом крупных партнеров, обеспечивая баланс между скоростью внедрения и технологической надежностью решений.

Рынок вендоров искусственного интеллекта в России

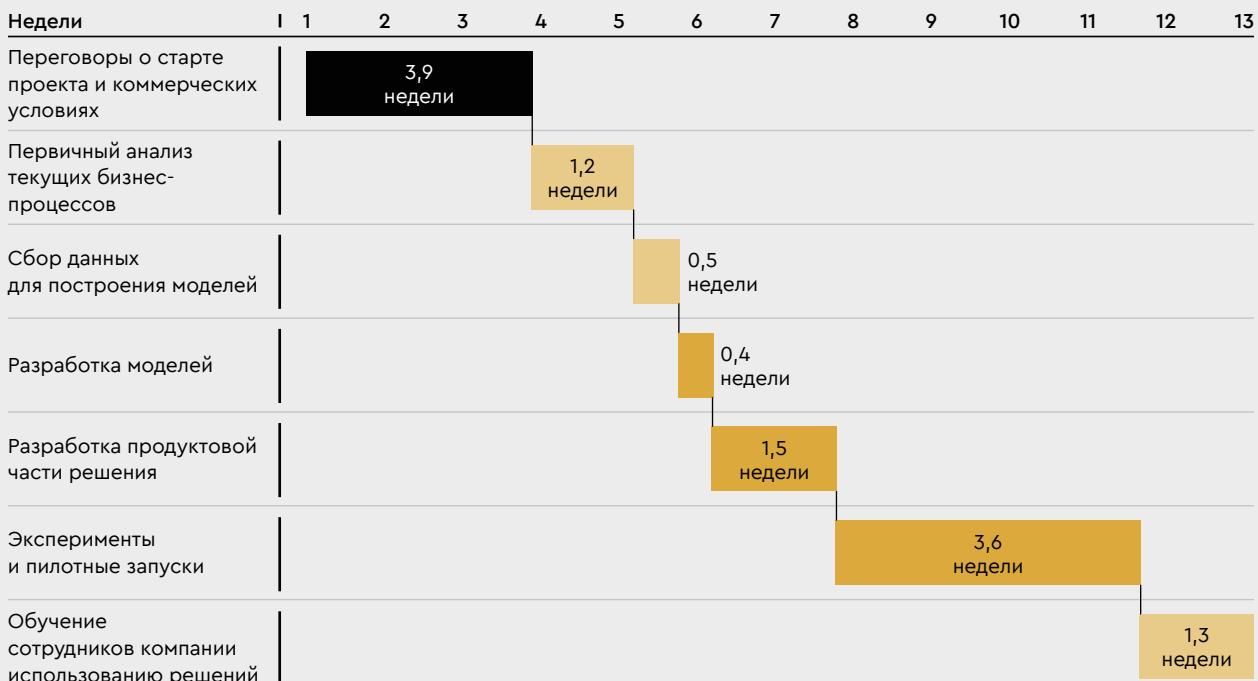
Российский рынок вендоров искусственного интеллекта является фрагментированным. Судя по результатам опроса вендоров, большинство вендоров специализируются на одной технологии — более 85% интеграторов ИИ-решений предоставляют решения только в одном направлении, концентрируясь на углубленной экспертизе и отработке конкретных сценариев применения. При этом на рынке присутствуют и крупные игроки, такие как «Яндекс» или «Сбер», которые покрывают полный спектр технологий — от генеративного ИИ и речевых решений до компьютерного зрения и инструментов машинного обучения.

Этапы реализации проектов по внедрению ИИ-решений

По результатам опроса вендоров, наиболее продолжительным этапом реализации проектов является подготовительный этап — переговоры о старте и согласование коммерческих условий, — который в среднем занимает около 4 недель. Техническая часть проекта, включая настройку и интеграцию решения, требует значительно меньше времени. Такая динамика объясняется тем, что большинство вендоров предлагают коробочные решения, не требующие длительной индивидуальной доработки, а ключевым фактором, влияющим на сроки, остаются процессы согласования и утверждения со стороны заказчика.

Средняя оценка вендорами длительности и сложности ключевых этапов проектов внедрения ИИ

Оценка сложности (1 — наименее, 5 — наиболее сложно) 1-3 3-4 4-5



Источник: опрос вендоров (блок 2. Продукт, вопрос 2.8. Какой из этапов реализации наименее предсказуем с точки зрения сроков и почему?)

С точки зрения сложности реализации вендоры отмечают, что наиболее трудоемким этапом являются именно переговоры о запуске проекта — средняя оценка сложности 4,6 из 5, что отражает необходимость детальной проработки требований и финансовых условий. При этом этот же этап был назван наименее предсказуемым — его указали все участники опроса как наиболее вариативный этап по срокам и результату. В то же время наиболее простым этапом в среднем оказалось обучение сотрудников заказчика (2 балла из 5), поскольку оно обычно проводится в упрощенном формате для ограниченной группы пользователей.

Модели сотрудничества вендоров в области ИИ

Вендоры также активно привлекают партнеров для оказания услуг и реализации проектов в сфере искусственного интеллекта, например для поставки оборудования, реализации дистрибуции решений или доработки продукта и основной технологии.

По результатам опроса вендоров, среди партнеров доминируют крупные технологические компании — с ними сотрудничают практически все участники рынка (99%), тогда как с другими специализированными ИИ-вендорами работают около 55%. Это отражает тенденцию к формированию экосистемных связей, где ключевые игроки объединяют усилия для масштабирования решений и выхода на новых клиентов.

По данным опроса, наибольшее вовлечение партнеров наблюдается на этапе дистрибуции решений для клиентов — его отметили 60% вендоров. При работе с инфраструктурой компании почти всегда используют аутсорсинг (95%), поскольку вендоры, как правило, сосредоточены на разработке программного обеспечения, а аппаратная часть — камеры, сенсоры и вычислительные мощности — передается внешним подрядчикам.

В то же время основная технология и продукт разрабатываются преимущественно внутри компаний при точечном подключении партнеров для решения специализированных задач. По результатам опроса вендоров, почти 90% компаний привлекают партнеров в формате консультирования или оказания отдельных услуг при разработке технологического ядра и почти 65% — при создании продуктовой части. Такой подход позволяет вендорам сохранять контроль над ключевыми компетенциями и интеллектуальной собственностью, минимизируя внешние зависимости.

Модели партнерств при разработке ИИ-решений



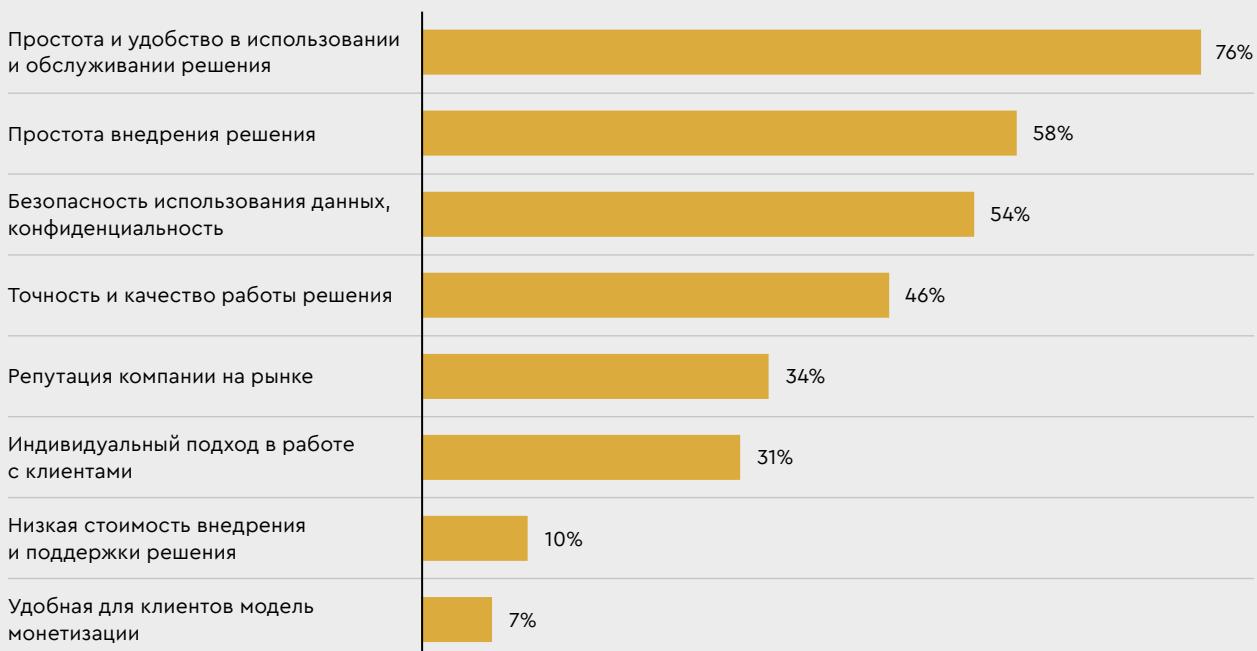
Источник: опрос вендоров (блок 2. Продукт, вопрос 2.3. Есть ли у компании партнеры, с которыми она работает при разработке и продаже решений?)

Модели сотрудничества вендоров в области ИИ

По результатам опроса 150 крупнейших российских вендоров, ключевые конкурентные преимущества компании видят не в ценовой политике, а в простоте и эффективности использования решений. Около 76% вендоров делают акцент на удобстве пользовательского опыта, 58% — на простоте внедрения в корпоративную инфраструктуру. На низкую стоимость решений ориентируется лишь 10% компаний, что связано с ограниченной долей малого и среднего бизнеса среди клиентов и фокусом на корпоративный сегмент.

Российский рынок искусственного интеллекта находится на этапе активного развития и структурного созревания, объединяя как устоявшиеся технологические решения, так и новые, быстро развивающиеся направления. Сегодня ИИ перестает быть нишевой областью и становится системным фактором конкурентоспособности бизнеса и государства.

В чем вендоры видят свои конкурентные преимущества, % ответа вендоров



Источник: опрос вендоров (блок 2. Продукт, вопрос 2.10. Как вам кажется, какие качества компании и продукта обеспечивают вам конкурентное преимущество и являются наиболее важными для клиентов?)

В ближайшие годы ключевыми драйверами его развития станут масштабирование генеративных моделей, внедрение речевых технологий и систем обработки естественного языка (NLP), совершенствование компьютерного зрения и развитие рекомендательных платформ. Особое значение также будет иметь синергия генеративного ИИ с другими технологиями. Эти направления формируют основу для ИИ-трансформации во всех ключевых отраслях.

Понимание текущих трендов, инвестиционной динамики и факторов роста рынка приобретает стратегическое значение: именно от способности участников экосистемы адаптироваться к новой технологической реальности будет зависеть темп экономического развития и позиция России на глобальной карте искусственного интеллекта. В дальнейших разделах мы детально расскажем о том, на каком уровне внедрения в компаниях разных отраслей находятся основные технологии ИИ и какие ключевые технологические изменения ожидаются в будущем.

Ожидаемый финансовый эффект от внедрения ИИ

Вклад искусственного интеллекта в экономику России будет расти. К 2030 г. искусственный интеллект перестанет быть инструментом точечной оптимизации и станет ключевым драйвером рентабельности и конкурентного преимущества в российской экономике.

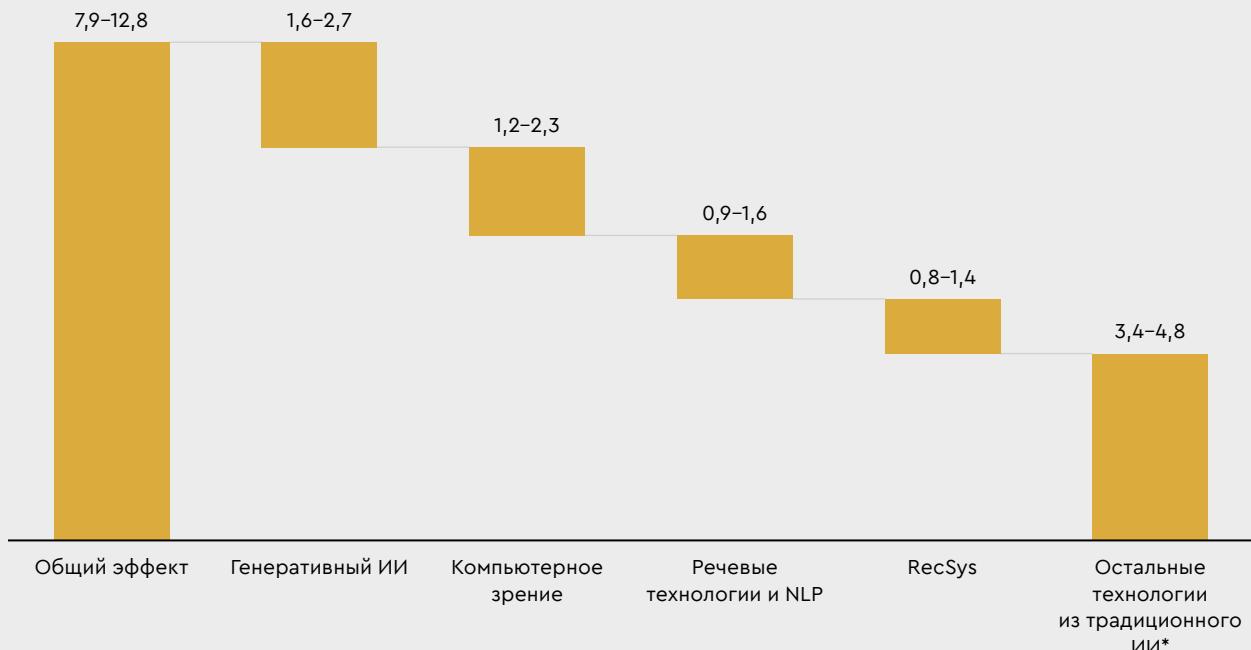
Мы провели количественную оценку вклада искусственного интеллекта в экономику России к 2030 г. Для расчета совокупного финансового эффекта использовалась модель, оценивающая влияние ИИ на операционную прибыль отраслей на основе нескольких факторов: прогнозируемой выручки 16 ключевых секторов экономики России, ожидаемых показателей рентабельности по отраслям на 2030 г. и оценки влияния технологий ИИ, основанной на отраслевых исследованиях и опросе более 150 компаний. Такой подход позволяет оценить как прямой экономический эффект от внедрения ИИ, так и его вклад в повышение операционной эффективности предприятий в различных отраслях.

Суммарный годовой финансовый эффект для российской экономики от внедрения технологий ИИ составит 7,9–12,8 трлн рублей к 2030 г. Это эквивалентно до 5,5% от прогнозируемого ВВП России в 2030 г.¹, или объему прибыли всей банковской отрасли. Обновленный прогноз превышает предыдущие оценки, которые составляли 4,2–6,9 трлн руб. и до 4% влияния на ВВП за 5-летний период до 2028 г. Рост связан со стремительным развитием технологий, в первую очередь генеративного ИИ, и их активным проникновением в бизнес-процессы. Важно подчеркнуть, что данный эффект представляет собой не просто экономию затрат, а создание новой стоимости через оптимизацию бизнес-процессов, создание новых продуктов и услуг, а также формирование принципиально новых бизнес-моделей.

Распределение эффекта по технологиям

Распределение эффекта по технологиям неравномерно. Среди ключевых технологий на рынке ИИ потенциальный эффект в будущем обратно пропорционален их текущей зрелости — генеративный ИИ и компьютерное зрение возглавляют ИИ-трансформацию.

Ожидаемый финансовый эффект от внедрения ИИ к 2030 г., трлн руб.



* Машинное обучение, глубокое обучение и продвинутая аналитика, исключая четыре ключевые технологии: генеративный ИИ, компьютерное зрение, речевые технологии и NLP, RecSys

Источник: анализ «Яков и Партнёры»

Анализ вклада отдельных технологических направлений показывает существенную дифференциацию.

Генеративный ИИ демонстрирует наиболее динамичный рост и высокий потенциал для трансформации отраслей, являясь лидером по вкладу в EBITDA среди рассматриваемых ИИ-технологий. При формировании прогноза в 2023 г. ожидаемый эффект от генеративного ИИ оценивался в диапазоне 0,8–1,3 трлн руб. На текущий момент наблюдается значительный рост потенциала технологии, обусловленный ее стремительным развитием и ускоренным внедрением в бизнес-процессы. Наибольший экономический эффект будет достигнут за счет автоматизации интеллектуального труда в таких областях, как создание контента, разработка продуктов и персонализированный маркетинг.

Компьютерное зрение сохраняет высокие позиции по вкладу в экономический эффект благодаря способности автоматизировать процессы контроля и анализа визуальной информации. Широкое внедрение решений на базе компьютерного зрения в производственные процессы останется ключевым драйвером роста.

Речевые технологии и NLP обеспечивают экономический эффект за счет автоматизации обработки текстовой и голосовой информации. Развитие технологий направлено на повышение эффективности взаимодействия с клиентами и оптимизацию работы с документами. При этом в сегменте создания контента и сложной аналитики текстов наблюдается активное замещение традиционных NLP-решений моделями генеративного ИИ, что приводит к перераспределению потенциала роста между технологиями.

Рекомендательные системы продолжат вносить значительный вклад в экономический эффект за счет повышения эффективности взаимодействия с потребителями и оптимизации ключевых бизнес-процессов. В таких направлениях, как персонализированный маркетинг и клиентские коммуникации, начинается интеграция с генеративными моделями для создания уникального контента под конкретного пользователя. При этом классические алгоритмы рекомендательных систем сохранят ведущую роль в решении базовых задач прогнозной аналитики и анализа пользовательского поведения.

Распределение эффекта по технологиям

Более 60% эффекта от внедрения ключевых технологий приходится на топ-5 отраслей. Крупнейшими бенефициарами внедрения ИИ в абсолютном выражении станут отрасли с высокими оборотами и рентабельностью.

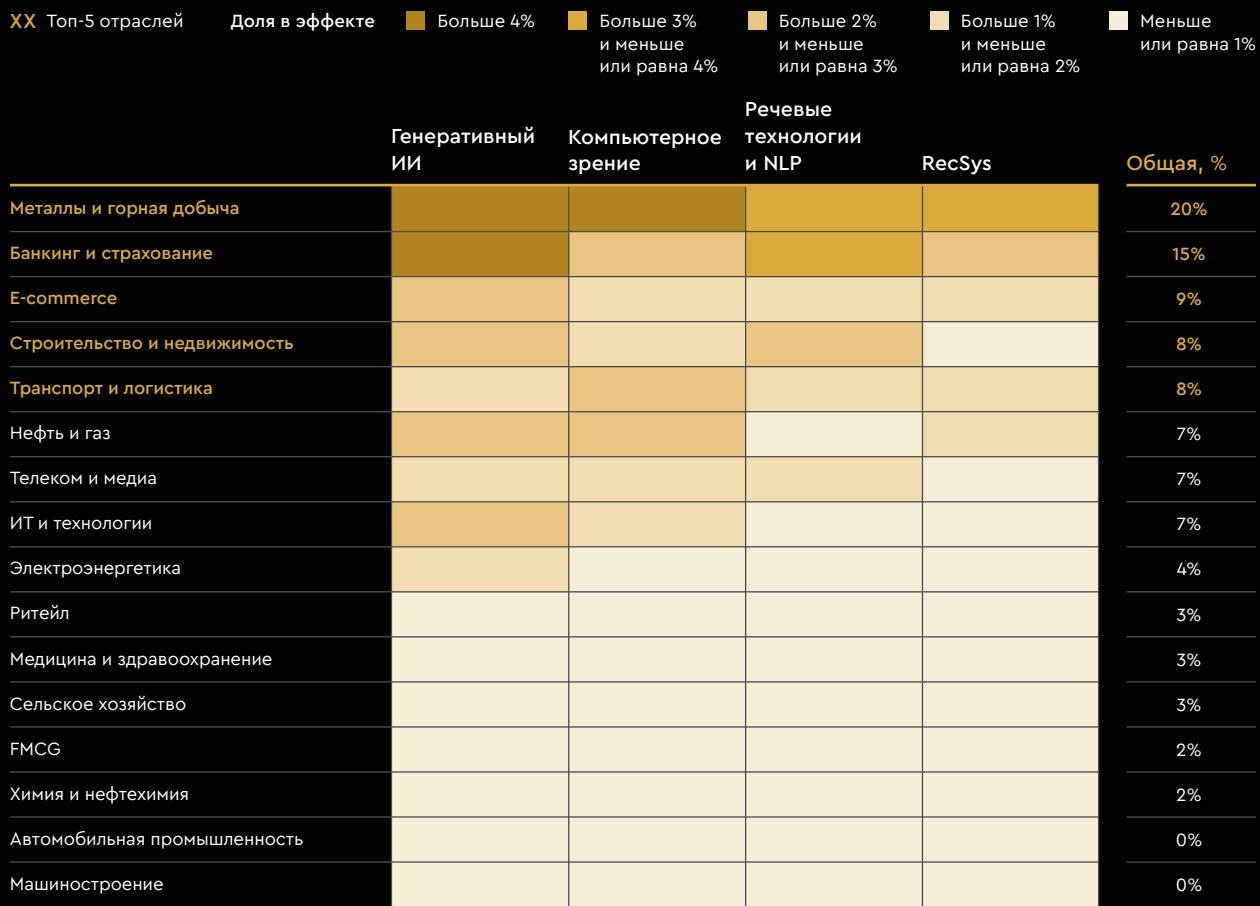
Наибольшую долю в общем эффекте займут отрасли, где внедрение ИИ-решений оказывает комплексное воздействие на операционную эффективность, клиентский опыт и создание новых продуктов.

Металлы и горная добыча. Ключевой вклад обеспечивает внедрение компьютерного зрения для автоматического контроля качества руды и безопасности работ, предиктивных моделей для обслуживания оборудования и ИИ-оптимизации логистических цепочек.

Банкинг и страхование. Генеративный ИИ трансформирует ключевые процессы, автоматизируя создание персонализированных предложений и условий продуктов. Технология дополняет существующие решения: NLP для обработки заявок и документов, компьютерное зрение для верификации клиентов и предиктивные модели для скоринга и оценки рисков.

E-commerce. Максимальный эффект достигается за счет комплексного внедрения технологий ИИ — создания уникального контента и описаний товаров на основе генеративного ИИ, формирования персонализированных рекомендаций через RecSys, оптимизации логистики с помощью предиктивных алгоритмов и автоматизации службы поддержки с использованием чат-ботов.

Доля отрасли в эффекте от внедрения 4 ключевых технологий к 2030 г., %



Источник: анализ «Яков и Партнёры»

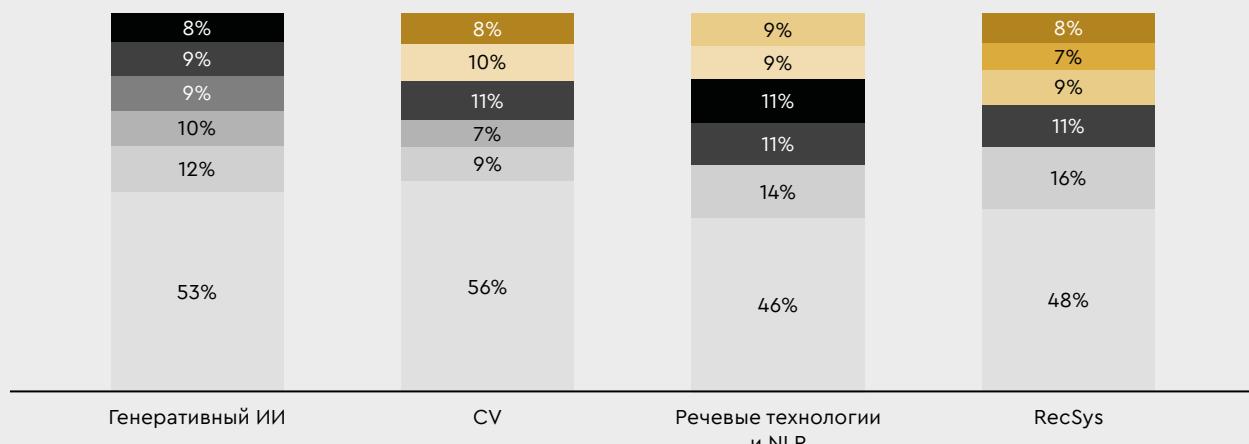
Строительство и недвижимость. Эффект складывается за счет автоматизации обработки клиентских обращений и документации, применения компьютерного зрения для мониторинга объектов и предиктивного анализа рынка недвижимости.

Транспорт и логистика. Основной вклад обеспечивают ИИ-оптимизация маршрутов, предиктивное обслуживание транспорта и автоматизация складских операций с использованием компьютерного зрения.

Удельный финансовый эффект на EBITDA позволяет устраниć влияние масштаба отраслей и выявить реальную глубину проникновения ИИ по отраслям. В отличие от абсолютных показателей, где доминируют крупнейшие по выручке сектора, данная метрика показывает, в каких отраслях каждый рубль операционной прибыли в наибольшей степени зависит от внедрения технологий искусственного интеллекта. Такой подход обеспечивает сопоставимость оценки между индустриями разного размера и демонстрирует стратегическую важность ИИ для сохранения конкурентных позиций бизнеса.

Доля топ-5 отраслей в удельном эффекте на EBITDA от внедрения технологий к 2030 г., %

Металлы и горная добыча Ритейл Банкинг и страхование Медицина и здравоохранение
Строительство и недвижимость Телеком и медиа Электроэнергетика ИТ и технологии
E-commerce Другое



Источник: анализ «Яков и Партнёры»

Удельная эффективность ИИ максимальна в отраслях, где технологии напрямую влияют на ключевые бизнес-процессы и монетизацию. Наибольший удельный эффект на EBITDA демонстрируют следующие пять секторов: e-commerce, телеком и медиа, ИТ и технологии, медицина и здравоохранение, а также строительство и недвижимость.

Если в абсолютных показателях лидируют традиционные отрасли с большими оборотами (металлы и горная добыча, банкинг и страхование), то в удельном эффекте доминируют сектора, где ИИ не просто оптимизирует процессы, а трансформирует саму бизнес-модель и создает новые источники монетизации. Это демонстрирует, что для digital-ориентированных отраслей внедрение ИИ становится критически важным для сохранения конкурентоспособности, тогда как в традиционных отраслях технологии в основном служат инструментом повышения операционной эффективности.

К 2030 г. внедрение ИИ станет вопросом выживания для большинства компаний. Распределение совокупного эффекта демонстрирует, что максимальную отдачу от инвестиций в искусственный интеллект получат сектора, где технологии оказывают непосредственное влияние на ключевые бизнес-процессы и модели монетизации. Компании, которые уже сегодня выстраивают стратегию вокруг данных и алгоритмов, закладывают фундамент для доминирования на своих рынках в следующем десятилетии.



Генеративный ИИ

Развитие технологии генеративного ИИ

2017 г.

- Заложен фундамент для архитектуры Transformer после выхода статьи Google под названием Attention Is All You Need («Внимание — это все, что вам нужно»), что в дальнейшем стало основой для создания предобученных языковых моделей GPT-1 и BERT.

2018 г.

- Выпуск GPT-1, первой версии трансформерной языковой модели от компании OpenAI, которая заложила основу для массового распространения языковых моделей.

2021–2022 гг.

- Генеративные модели научились работать с изображениями — прорывом в этом направлении стала OpenAI с моделью DALL·E, а также Midjourney.
- Началось развитие open-source-моделей, появилась YaLM 100B от «Яндекса».

2023 г.

- Модели перешли от генерации текстов к выполнению конкретных инструкций.
- Появление российских LLM, таких как YandexGPT от «Яндекса» и GigaChat от «Сбера», Perplexity.
- Заложена основа для развития ИИ-агентов благодаря моделям, которые научились обращаться к внутренним и внешним источникам данных перед генерацией ответа.
- Началось распространение open-source-моделей (LLaMA, DeepSeek) и датасетов, которые обеспечивают конкурентное качество при меньших затратах и легко адаптируются под отраслевые задачи и существующую цифровую инфраструктуру.

2024 г.

- Модели начали выполнять действия в реальном мире по текстовым запросам, расширяя возможности робототехники.
- Модели начали активно внедряться в бизнес-процессы, автоматизируя задачи и улучшая качество продуктов.
- Модели начали активно интегрироваться в поисковые системы: Google запустил MUM для обработки сложных запросов, «Яндекс» представил Neural Search для улучшения поиска с учетом смысла запросов, а ChatGPT получил возможность искать информацию в интернете в реальном времени.
- В мультимодальных системах появились голосовые режимы.
- Увеличилось качество генерации видео — ролики стали длиннее и реалистичнее.
- Увеличение контекстного окна позволило моделям учитывать длинную историю запросов и работать с большими объемами данных.
- Модели ИИ делают значительный шаг от простого создания контента к способности рассуждать, выстраивать логические цепочки и планировать действия.

2025 г.

- Появились универсальные ИИ-агенты, способные самостоятельно подбирать инструменты и выполнять многошаговые задачи, а на платформах, таких как GPTees, Yandex AI Studio, Google AI Studio, появилась возможность создавать и настраивать собственных агентов.
- Фокус в области ИИ стал смещаться от создания моделей с нуля к запуску прикладных решений, основанных на open-source-моделях.
- По мере того как ИИ становится все более мощным и автономным, усиливается фокус на его ответственном и этичном использовании.
- Проводятся активные работы по оптимизации и квантованию моделей, направленные на повышение их эффективности, в ответ на неконтролируемый рост числа параметров и связанных с этим вычислительных затрат.

Текущее положение России в развитии генеративного ИИ

Степень внедрения генеративного ИИ в российских компаниях за последние три года значительно продвинулась. Согласно опросу СТО, технология прошла путь от экспериментов к массовому внедрению за 2 года.

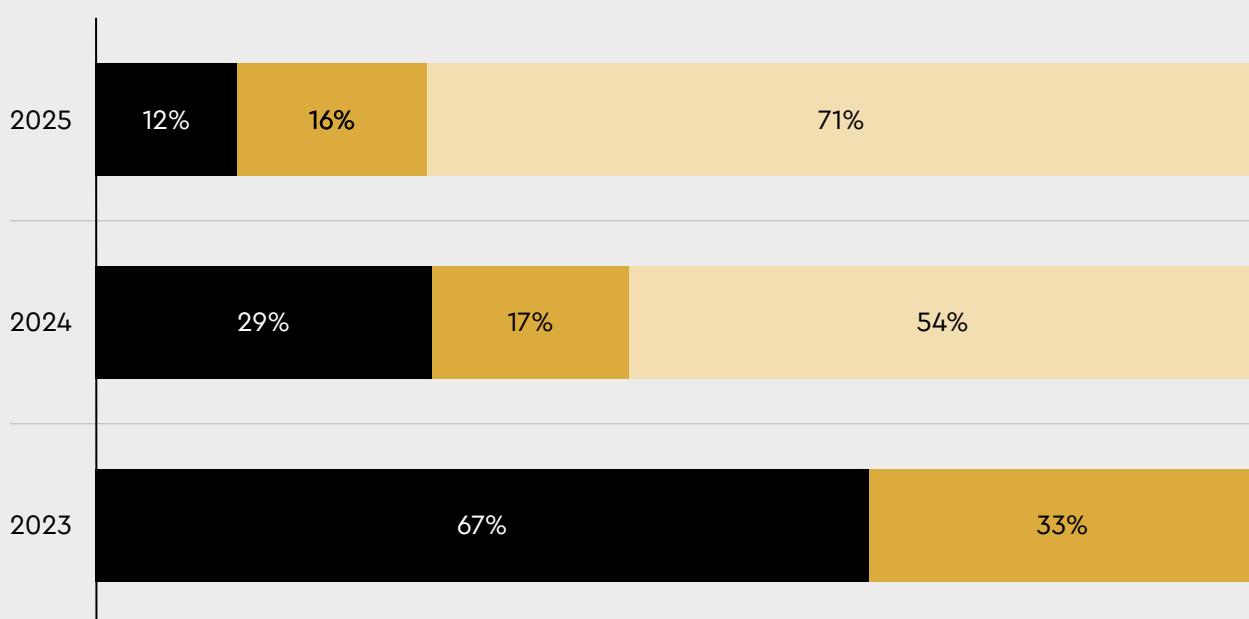
Если в 2023 г. полноценного внедрения генеративного ИИ в российских компаниях не фиксировалось вовсе, то уже в 2024 г. более половины организаций сообщили о применении генеративного ИИ хотя бы в одной функции бизнеса, а к 2025 г. этот показатель достиг 71%, что на 17 п. п. выше, чем годом ранее. Всего за два года технология прошла путь от единичных pilotных инициатив к массовому внедрению, продемонстрировав один из самых стремительных темпов распространения среди всех направлений искусственного интеллекта.

Тем не менее, несмотря на бурное развитие технологии, она еще на пути полной интеграции в повседневную жизнь пользователей: с концепцией генеративного ИИ знакома практически половина потребителей, но только 26% россиян старше 12 лет хотя бы раз в месяц используют какой-либо сервис.

В ТОП-5 самых популярных сервисов у россиян вошли: Алиса AI (проникновение 14%), DeepSeek (9%), GigaChat (4%), ChatGPT (4%), Perplexity AI (1%)²

Степень внедрения генеративного ИИ, % компаний

■ Проводятся эксперименты по работе с генеративным ИИ в свободном режиме ■ Определены приоритетные функции и бизнес-кейсы внедрения генеративного ИИ, решение не внедрено ■ Решения на основе генеративного ИИ полноценно внедрены хотя бы в одной функции организации



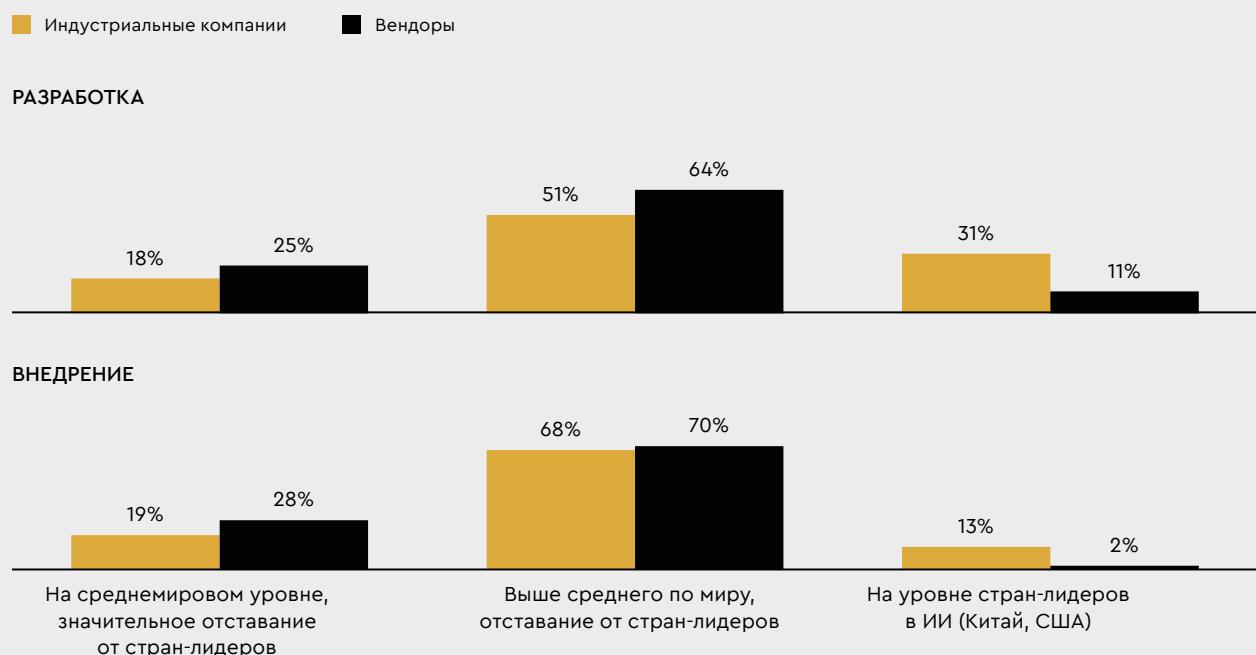
Источник: опрос индустриальных компаний России, проведенный «Яков и Партнёры» в 2023, 2024 и 2025 гг.

При этом качество как разработки, так и внедрения генеративного ИИ в России оценивается высоко. Согласно опросу СТО, почти треть компаний (31%) считают уровень разработки генеративного ИИ в стране сопоставимым с мировыми лидерами, а 13% полагают, что и по степени внедрения Россия уже находится на их уровне. Это подтверждается тем, что уже сейчас в России существуют конкуренты крупнейшим LLM-моделям стран-лидеров, таким как DeepSeek и ChatGPT, например Alice AI (предыдущее поколение нейросети называлось YandexGPT) и GigaChat.

Отдельно стоит отметить, что половина пользователей также уверены, что Россия занимает позицию «выше среднего по миру» в области разработки и внедрения генеративного ИИ.

Тем не менее большинство участников опросов со стороны вендоров и индустриальных компаний — свыше 50% по разработке и около 70% по внедрению — отмечают, что Россия находится «выше среднего по миру», но с некоторым отставанием от стран-лидеров. Это подтверждает, что отечественная экосистема генеративного ИИ уже сформировалась и воспринимается как сильная и перспективная. При этом, по мнению и компаний, и вендоров, разработка в России опережает внедрение, что указывает на наличие прочной технологической базы, потенциал которой еще не полностью реализован в практических бизнес-кейсах.

Восприятие уровня разработки и внедрения генеративного ИИ в России, % компаний и вендоров



Источник: опрос индустриальных компаний (блок 1. Контекст, вопрос 1.1. Как вы оцениваете текущий уровень разработки и внедрения технологий ИИ в России?); опрос вендоров (блок 1. Контекст, вопрос 1.1. Как вы оцениваете текущий уровень разработки и внедрения технологий ИИ в России?)

Сегодня генеративный искусственный интеллект становится не только технологическим, но и важным внешне- и внутриполитическим фактором. Страны осознают стратегическую важность этой технологии и стремятся развивать собственные решения на базе генеративного ИИ, чтобы не попасть в зависимость от внешних игроков. Россия, обладая необходимым потенциалом и компетенциями, проявляет явный запрос на технологическую независимость и активно работает над созданием собственных решений, не зависящих от стран-лидеров.

Ключевые тренды и изменения в технологии генеративного ИИ

В 2025 г. развитие генеративного искусственного интеллекта вступило в новую фазу — от масштабирования моделей к повышению их автономности, независимости и мобильности. Ключевые изменения затронули архитектуру ИИ-систем, способы их интеграции с корпоративной инфраструктурой и данными: появилось больше моделей, способных рассуждать, работать локально и безопасно взаимодействовать с корпоративной инфраструктурой. Существенную роль в этом сыграли развитие открытых экосистем, стандартизация протоколов обмена данными и рост вычислительных возможностей пользовательских устройств. На этом фоне сформировались ключевые технологические тренды 2025 г., определяющие дальнейшую эволюцию генеративного ИИ.

Мультимодальные модели

Новый класс моделей, способных одновременно воспринимать и обрабатывать разные типы данных — текст, изображения, аудио, видео и другие форматы. Благодаря единому представлению информации такие системы могут понимать контекст комплексно: описывать картинки словами, анализировать документы со схемами, интерпретировать видео или звук вместе с текстом. Ключевым трендом в развитии мультимодальности становится переход к end-to-end-омнимоделям, способным работать с любыми типами данных в едином пространстве представлений (например, Qwen3-Omni).

Рассуждающие модели (Reasoning LLM)

Новый режим работы моделей, при котором они используют больше вычислений, чтобы шаг за шагом продумывать решение и проверять его правильность перед тем, как дать ответ. Вместо того чтобы увеличивать размер модели, рост качества достигается за счет более глубокого процесса размышления. Это дает заметный прирост точности в написании кода, решении математических задач и выполнении логических рассуждений.

В январе 2025 г. компания DeepSeek AI представила модель DeepSeek-R1 с «встроенным рассуждением» (reasoning) — способностью генерировать цепочки рассуждений (chain-of-thought) для сложных задач по математике, коду и логике. В России существует аналогичная функция (например, в моделях от «Яндекса» и «Сбера»).

Развитие и повышение значимости open-source-моделей

Большинство компаний в 2025 г. используют open-source-модели, поскольку отказ от них экономически и технологически невыгоден — это замедляет разработку и выводит из технологической гонки, при этом практически все современные ИИ-системы в той или иной степени включают open-source-компоненты. Использование открытых моделей ускоряет вывод решений на рынок, снижает барьеры для бизнеса и способствует массовому внедрению ИИ. Особую роль этот подход играет в условиях ограниченного доступа к вычислительным ресурсам и санкционного давления — поэтому дообучение open-source-моделей особенно востребовано в России.

Существуют два основных подхода к адаптации ИИ-моделей: инициализация с использованием предобученных весов и дообучение моделей.

В первом варианте компании проводят полный цикл обучения моделей, контролируя все этапы: сначала берут предобученные «веса» (базовые настройки модели), чтобы сэкономить ресурсы на старте, затем выполняют собственный претрейн — полностью переобучают модель на своих данных. При необходимости на этапе претрейна может быть изменена архитектура модели (например, добавлены новые слои или скорректированы параметры). Далее следует донастройка (алайнмент), в ходе которой модель адаптируется к конкретным задачам: уточняется логика ответов, корректируется стиль, добавляются специфические функции. Также в процесс может быть включён RL (Reinforcement Learning, обучение с подкреплением) — метод, при котором модель «учится» через пробы и ошибки: за правильные действия она получает «награду», за ошибки — «штраф», постепенно улучшая качество своих ответов. Порядок и сочетание этих этапов (алайнмента и RL) зависят от поставленной задачи и выбранной стратегии обучения. Такой подход требует огромных вычислительных мощностей, команды высококвалифицированных специалистов и развитой инфраструктуры — по сложности он сопоставим с обучением модели с нуля, поскольку в итоге модель кардинально меняется даже при старте с предобученных весов. Разные подходы к гибридной инициализации — сочетании open-source-или предобученных компонентов с глубокой доработкой внутри компании — активно применяется мировыми лидерами, такими как Meta, Microsoft, Google, а в России — Яндексом.

Второй подход — дообучение — более доступный и практичный. Дообучением принято называть самую последнюю и наименее вычислительно затратную фазу (Supervised Fine-Tuning). Она заключается в том, чтобы взять уже готовую и хорошо работающую модель и «подтянуть» её под конкретные бизнес-задачи: научить лучше понимать корпоративные данные, стиль общения или специфику отрасли. Такой способ менее затратен, что делает его самым популярным среди многих вендеров.

Развитие агентских систем

Направление, связанное с созданием автономных ИИ, способных выполнять задачи без постоянного вмешательства человека. В отличие от традиционных моделей, работающих в формате «вопрос — ответ», такие агенты умеют планировать действия, использовать внешние инструменты и взаимодействовать с бизнес-системами. Это существенно расширяет спектр практических сценариев — от автоматизации внутренних процессов до разработки новых клиентских сервисов. Для компаний это означает переход от использования ИИ в качестве помощника к модели, где агент становится полноценным участником бизнес-операций. Примеры таких агентов включают AutoGPT, Devin, Adept, AI Agents от Cognition, персонажей в Character.AI или Inworld, а также пользовательских агентов в Alice AI и браузере.

Платформы для разработки ИИ-агентов

Качество конечных бизнес-решений на базе генеративного ИИ обеспечивают не только сами модели, но и инструменты для их адаптации под конкретные задачи бизнеса. С их помощью можно настроить генерацию ответа по интернету и базам знаний, создать голосового ИИ-агента или настроить логику взаимодействия ИИ-агента и внешних систем.

Для создания ИИ-агентов активно развиваются платформы, такие как Microsoft Azure AI Studio, Google Vertex AI Agent Builder, Meta AI Studio, OpenAI GPTs, Rasa, Botpress, aitaro.ai. В России также активно развиваются платформы для создания ИИ-агентов, в том числе Yandex AI Studio.

Генеративный ИИ на устройствах

Новый подход, при котором генеративные модели стало возможно запускать прямо на ноутбуках и смартфонах благодаря росту мощности нейропроцессоров (NPU). Если раньше такие вычисления были возможны только с использованием облачных мощностей, то теперь они работают непосредственно на устройствах, что уменьшает задержку и помогает обрабатывать данные, требующие конфиденциальности. Так, компания Apple в 2025 г. представила чип M3, который позволяет запускать генеративные модели прямо на устройствах, что значительно ускоряет работу с ИИ и обеспечивает высокий уровень конфиденциальности для пользователей.

Аналогичные исследования активно развиваются и в России, где квантование моделей также становится важной частью научных разработок.

Происхождение контента

Новый стандарт, при котором к сгенерированным материалам автоматически добавляются метки происхождения, фиксирующие, где и каким образом был создан или изменен контент. Это помогает бороться с подделками и становится все более важным требованием для использования технологий в медиа, корпоративной среде и государственных сервисах. Так, в феврале 2025 г. Google объявил, что фотографии, отредактированные с помощью функции Magic Editor AI, будут автоматически помечаться цифровыми водяными знаками SynthID. В России аналогичные решения уже тестируются в нейросетях YandexART и на платформе Kandinsky от «Сбера».

Оптимизация стоимости инференса

Компании фокусируются на снижении стоимости инференса (процесса обработки данных моделями), улучшая алгоритмы и повышая энергоэффективность моделей. Это позволяет быстрее и дешевле внедрять генеративный ИИ в массовом масштабе, расширяя его применение в различных секторах. Например, процессор NVIDIA A100 значительно улучшил энергоэффективность инференса для глубоких нейронных сетей, ускорив время обработки и снизив затраты на вычисления, что позволяет компаниям оптимизировать ресурсы и ускорить внедрение ИИ.

Mixture of Experts (MoE)

Архитектурный подход, в котором вместо наращивания одной «универсальной» модели формируется пул специализированных экспертов — небольших нейросетевых подсетей (мини-моделей), каждая из которых обучена на узком типе данных или подзадаче. Каждый эксперт обучается на своем подмножестве данных или аспекте задачи, а маршрутизатор (router) для каждого входа выбирает небольшое подмножество экспертов, которые действительно нужны именно сейчас. Затем механизм объединения агрегирует их ответы в финальный результат. Такой подход применяется в Alice AI, где используется MoE-архитектура.

Переход ИИ в физический мир

Интеграция ИИ в роботов и носимые устройства (например, носимые датчики, наушники или диктофоны) позволяет системам воспринимать окружение, принимать решения и выполнять действия в реальности. В кейсах Tesla Optimus и 1X NEO генеративные модели используются прежде всего как языковой интерфейс — для понимания команд, диалога и интерпретации инструкций. Обучение и исполнение самих действий (сортировка, перенос, уборка и т. п.) обеспечивают другие компоненты стека: компьютерное зрение, поведенческое клонирование/обучение с подкреплением, планирование и низкоуровневые контроллеры.

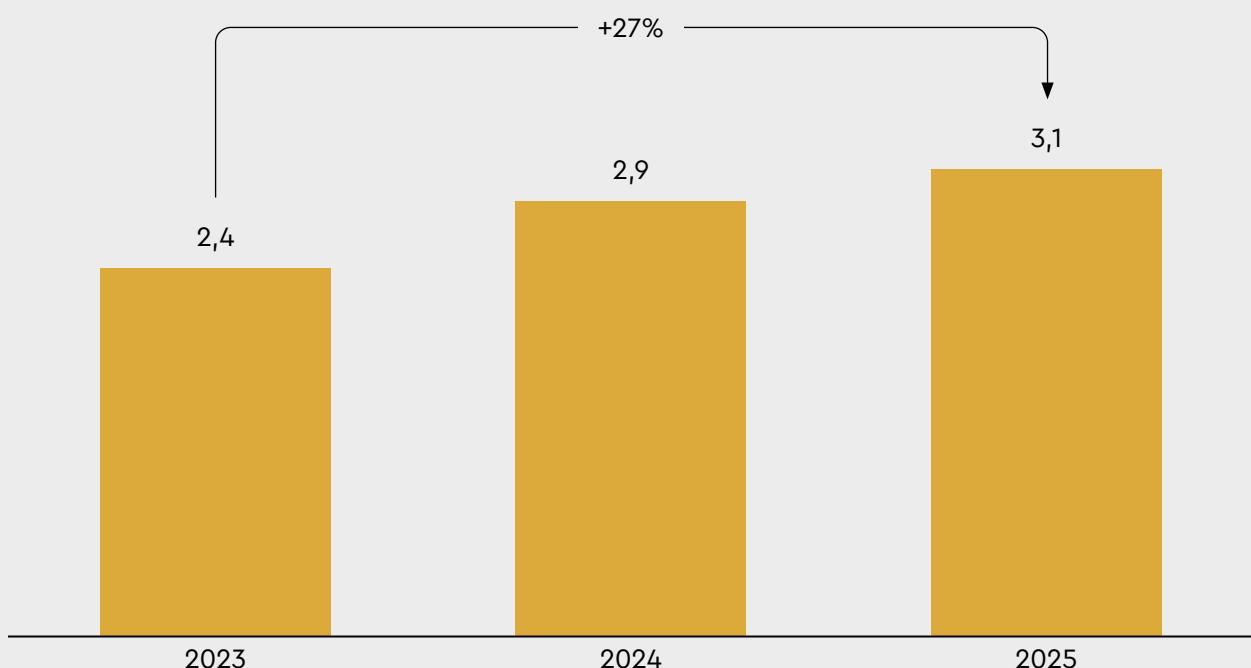
Основные области внедрения технологии генеративного ИИ

Широта применения генеративного ИИ

Генеративный искусственный интеллект — одна из самых активно развивающихся технологий в корпоративном сегменте. Компании все шире применяют его возможности в различных бизнес-функциях, увеличивая количество как пилотных, так и внедренных решений.

По результатам исследования, количество функций, где запущены пилоты или реализовано внедрение решений на базе генеративного ИИ, продолжает расти ежегодно: с 2023 г. показатель увеличился на 27% — с 2,4 до 3,1 функции в 2025 г. Это подтверждает закрепление генеративного ИИ как одного из ключевых инструментов цифровой трансформации бизнеса.

Среднее число функций, в которых реализованы внедрение или пилот, ед.



Источник: опрос индустриальных компаний России, проведенный «Яков и Партнёры» в 2023, 2024 и 2025 гг.

Универсальность генеративного ИИ отражается в его широкой применимости — технология используется уже в 80% ключевых бизнес-функций. При этом глубина внедрения существенно варьируется в зависимости от уровня технологической зрелости отраслей.

В среднем решения на базе генеративного ИИ уже внедрены (без учета пилотов) примерно в 3 функциях бизнеса, однако в продвинутых по уровню развития ИИ секторах этот показатель значительно выше — 3,9 функции против 2,4 в прочих отраслях. Лидерами по масштабу внедрения выступают ИТ (5,5 функции), e-commerce (4,1) и строительство (3,5), где генеративный ИИ уже стал частью ключевых бизнес-процессов и активно влияет на производительность.

Среднее количество функций, в которые внедряли генеративный ИИ, ед.

Продвинутые в ИИ отрасли



Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.1. В каких функциях внедряется данная технология и в какой степени?)

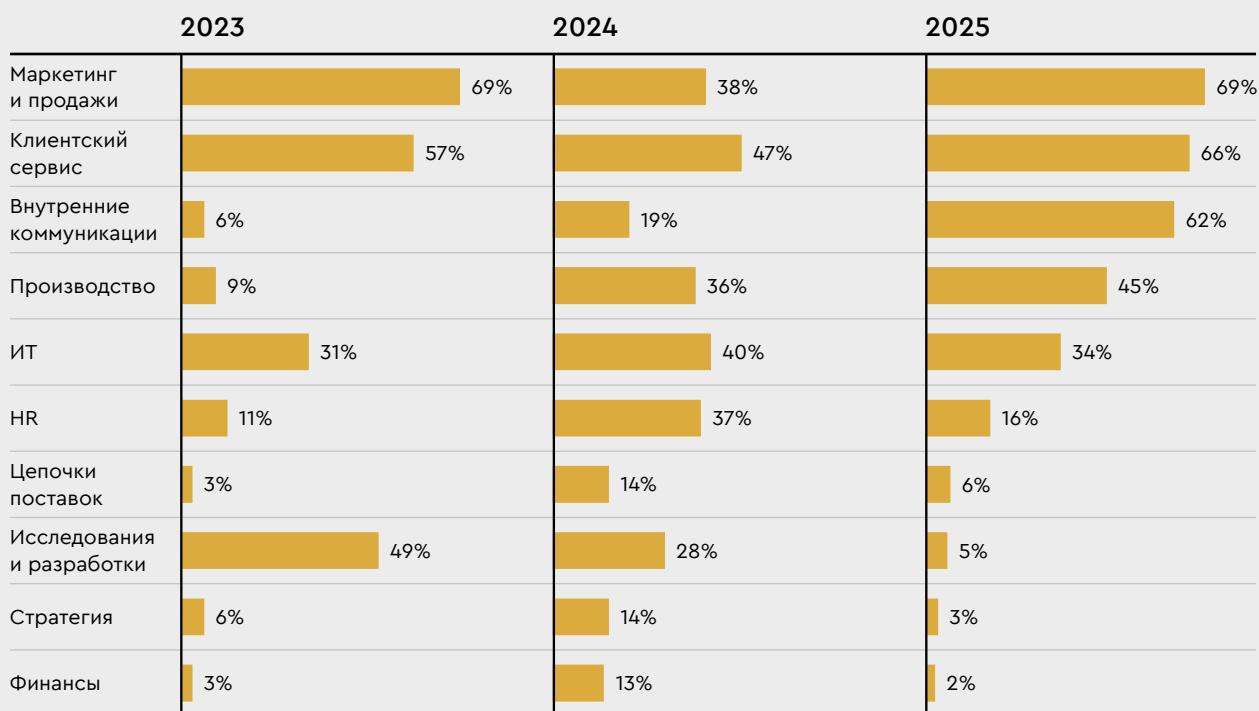
При этом, согласно опросу населения России, генеративный ИИ наиболее активно применяется в ИТ, телекоммуникациях, e-commerce, FMCG и ритейле.

Основные функции и сценарии внедрения

Динамика последних трех лет показывает эволюцию подхода компаний к использованию генеративного искусственного интеллекта. В 2023 г. внедрение и пилотирование генеративного ИИ было сосредоточено в наиболее понятных и прикладных функциях — маркетинге и клиентском сервисе. В 2024 г. фокус временно сместился: компании начали активнее осваивать менее традиционные области — HR, производство, цепочки поставок и финансы, что отражало поиск новых направлений применения технологии.

В 2025 г., по результатам опроса СТО, лидерами по внедрению и пилотированию решений на базе генеративного ИИ стали маркетинг и продажи (их отметили 69% компаний, уже внедривших или тестирующих технологию), клиентский сервис (66%) и внутренние коммуникации (62%). Особенno заметен рост интереса к внутренним коммуникациям: их доля увеличилась с 6% в 2023 г. до 62% в 2025 г., поскольку компании все чаще используют генеративный ИИ для автоматизации документооборота, подготовки корпоративных сообщений и оптимизации административных процессов. Такая динамика показывает, что компании вновь сосредоточились на клиентоориентированных функциях, а также на сценариях, где технология оказывает наибольшее влияние на операционную эффективность.

Функции, в которых внедрялись или пилотировались решения на базе генеративного ИИ, % компаний



Источник: опрос индустриальных компаний России, проведенный «Яков и Партнёры» в 2023, 2024 и 2025 гг.

По результатам опроса СТО, процент компаний, реализовавших полное внедрение генеративного ИИ, пока остается ниже уровня компаний, которые провели хотя бы один pilotный проект, однако структура лидирующих направлений сохраняется прежней.

Наиболее активно технологии генеративного ИИ применяются во внутренних коммуникациях (69%) — прежде всего для автоматизации обработки нормативных, юридических и регламентных документов; в маркетинге и продажах (63%), где ключевыми сценариями стали генерация маркетингового контента для рекламы и коммуникаций и цифровой виртуальный ассистент для взаимодействия с клиентами; а также в клиентском сервисе (48%), где чаще всего реализуются решения по автоматизированному чат-боту клиентской поддержки и анализу отзывов клиентов.

Эти направления демонстрируют наибольший эффект от применения генеративного ИИ, одновременно повышая качество клиентского опыта, скорость внутренних процессов и общую операционную эффективность бизнеса.

Генеративный ИИ часто комбинируется с другими технологиями, такими как NLP и речевые технологии, CV и RecSys, что позволяет создавать более мощные и комплексные решения. По результатам опроса были выделены топ-5 наиболее распространенных кейсов применения генеративного ИИ.



Доля компаний, внедривших генеративный ИИ в бизнес-функции, от общего числа опрошенных компаний по индустрии, %

	91-100%	61-90%	31-60%	1-30%	Продвинутые в ИИ отрасли	ИТ	Цепочки поставок	Исследования и разработки	Стратегия	Финансы	Среднее кол-во функций
ИТ и технологии	100%	100%	100%	89%	75%	90%					5,5
E-commerce	40%	100%	75%	25%		67%	67%	33%			4,1
Строительство и недвижимость	88%	90%	92%	50%	33%						3,5
Медицина и здравоохранение	100%	100%	90%	50%							3,4
Банкинг и страхование	78%	79%	100%		25%	57%					3,4
Электроэнергетика	100%	100%	50%	75%							3,3
Сельское хозяйство	100%	75%	50%	25%	50%						3,0
Металлы и горная добыча		33%	84%	56%	25%	20%	50%				2,7
Телеком и медиа	50%	100%		40%		67%					2,6
FMCG	100%	100%		50%							2,5
Транспорт и логистика	71%	50%		20%	100%						2,4
Автомобильная промышленность	100%		71%	67%							2,4
Ритейл	50%	76%	40%			38%	14%				2,2
Химия и нефтехимия	100%			60%	50%						2,1
Машиностроение	25%		20%	43%							0,9
Нефть и газ				25%	50%						0,8
Продвинутые в ИИ	67%	95%	69%	38%	25%	70%	17%	8%	0%	0%	3,9
Обычные отрасли	69%	52%	41%	43%	26%	5%	5%	0%	0%	0%	2,4
В среднем	69%	63%	48%	42%	26%	21%	8%	2%	0%	0%	2,8

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологий, вопрос 2.1. В каких функциях внедряется данная технология и в какой степени?)

Топ-5 бизнес-кейсов использования технологии генеративного ИИ (1/2)

Топ-1 бизнес-кейс по популярности в использовании

Генеративный ИИ

Генерация маркетингового контента для рекламы и коммуникаций (маркетинг и продажи)

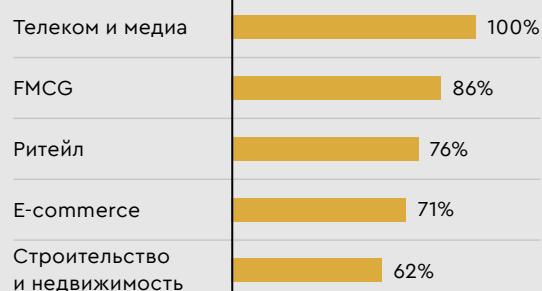
Описание

ИИ автоматически создает тексты, визуалы и сценарии для рекламы и коммуникаций, учитывая стиль бренда и целевую аудиторию. Система интегрируется с маркетинговыми платформами и ускоряет запуск кампаний

Результат

Быстрое создание контента и рост эффективности маркетинга

Частота использования*, %



Топ-2 бизнес-кейс по популярности в использовании

Генеративный ИИ

Автоматизация обработки внутренних нормативных, юридических и регламентных документов (внутренние коммуникации)

Описание

Помогает сотрудникам быстро находить и использовать документы, автоматически анализируя и классифицируя их, извлекая ключевую информацию. Система упрощает обмен информацией между подразделениями, интегрируясь с внутренними хранилищами и системами документооборота

Результат

Сокращение времени на обработку документов, повышение прозрачности взаимодействия и снижение рисков ошибок

Частота использования*, %



Топ-3 бизнес-кейс по популярности в использовании

Генеративный ИИ, речевые технологии и NLP

Цифровой виртуальный ассистент для взаимодействия с клиентами (маркетинг и продажи)

Описание

ИИ-ассистент сопровождает клиента на всем пути покупки — консультирует по продуктам, выполняет функции перекрестных продаж. Система обучается на данных о поведении клиентов и CRM, обеспечивая персонализированные рекомендации и повышение конверсии продаж

Результат

Повышение качества обслуживания и рост удовлетворенности клиентов

Частота использования*, %



* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

Топ-5 бизнес-кейсов использования технологии генеративного ИИ (2/2)

Топ-4 бизнес-кейс по популярности в использовании

Генеративный ИИ, речевые технологии и NLP, RecSys

Автоматизированный чат-бот клиентской поддержки (клиентский сервис)

Описание

Чат-бот на основе ИИ принимает обращения клиентов, отвечает на типовые запросы, помогает решать распространенные проблемы. Система интегрируется с базами знаний, CRM и внутренними системами компании, обеспечивая быструю и точную обработку запросов без участия оператора

Результат

Сокращение нагрузки на поддержку и повышение скорости обслуживания

Частота использования*, %

E-commerce	86%
Банкинг и страхование	60%
Телеком и медиа	60%
Ритейл	59%
ИТ и технологии	50%

Топ-5 бизнес-кейс по популярности в использовании

Генеративный ИИ, речевые технологии и NLP

Анализ отзывов клиентов (клиентский сервис)

Описание

Система с помощью ИИ анализирует тексты отзывов и обращений, выявляет темы, эмоции и причины недовольства. Отчеты формируются автоматически и помогают выявлять проблемные зоны и точки роста сервиса

Результат

Повышение качества обслуживания и лояльности клиентов

Частота использования*, %

Банкинг и страхование	60%
E-commerce	57%
Строительство и недвижимость	54%
Транспорт и логистика	50%
Телеком и медиа	50%

* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

Бизнес-кейсы применения генеративного ИИ продолжают развиваться, и мы ожидаем, что в будущем они могут значительно расширяться. Например, значимо разовьется гиперперсонализация, которая позволит создавать контент, ориентированный на каждого пользователя индивидуально. Это может включать персонализацию видео, рекламы и других взаимодействий на уровне предпочтений конкретного человека, что позволит компаниям создавать максимально релевантные предложения и улучшать клиентский опыт.

Роль ИИ-агентов в развитии технологий генеративного ИИ

Несмотря на активное развитие ИИ-агентов в 2025 г., сегодня сохраняется низкий уровень понимания различий между ИИ-агентами и ИИ-помощниками, а также их функционального разграничения. Этот факт подтверждается результатами как опроса технологических вендоров, так и исследований пользовательской аудитории.

По результатам опроса российских разработчиков и интеграторов ИИ-решений, разработчики ИИ-помощников в широком понимании — самое популярное направление: 34% вендоров предлагают продукты этого типа. Однако в большинстве случаев речь идет о B2B-решениях — поисковых RAG-агентах, инструментах автоматизации внутренних процессов и поддержке корпоративных задач, что отражает прикладной, прагматичный характер развития этого направления в России.

Аналогичная ситуация наблюдается и среди населения России: только 11% респондентов хорошо разбираются в концепции «ИИ-агент», при этом 58% имеют очень ограниченное представление о том, что это такое. При этом те, кто знаком с этим понятием, часто воспринимают его как более широкий термин, включая в эту категорию в основном чат-ботов и помощников для личных задач.

ИИ-агенты в сравнении с простыми ИИ-помощниками имеют следующие различия:

- Автономность: ИИ-агенты могут выполнять задачи без постоянного вмешательства человека.
- ИИ-ядро (модель): ИИ-агенты могут использовать модели, необязательно являющиеся LLM, что позволяет им быть более специализированными.
- Доступ к внешним инструментам: ИИ-агенты имеют возможность взаимодействовать с внешними системами, такими как базы данных, API и другие сервисы, что делает их более гибкими и функциональными.
- Гибкость и адаптивность: в отличие от предсказуемых ассистентов, ИИ-агенты могут изменять свои действия в зависимости от контекста и текущих задач, что позволяет им эффективно работать в динамичных условиях.
- Планирование и память: умение разбивать цель на шаги, сохранять контекст и результаты для последующих действий.
- Наблюдаемость и безопасность: самопроверка, логирование действий, политики и лимиты на вызовы инструментов.

Основные компоненты ИИ-агентов

Перцепция	Сенсоры, API, пользовательский ввод, документы, веб	Обработка естественного языка структурированных данных
Память	Краткосрочная (контекст текущей сессии)	Долгосрочная (база знаний, накопленный опыт, история взаимодействия)
Планирование	Постановка задач	Разработка плана действий
Действие	Выполнение задачи (веб-действия, скрипты, отправка писем и т. п.)	Взаимодействие с другими ИИ-агентами, моделями и системами
Рефлексия	Самооценка, корректировка планов, обучение на ошибках и т. п.	Рассуждение и принятие решений

Источник: анализ «Яков и Партнёры»

Ключевая ценность ИИ-агентов заключается в автоматизации сложных процессов «под ключ» — без необходимости ручного управления и постоянного участия человека. Они уже находят широкое применение в различных отраслях: от разработки кода и работы с документами до анализа данных, взаимодействия с BI-инструментами и выполнения рутинных задач бэк-офиса. В ближайшее время ИИ-агенты войдут и в повседневную жизнь пользователей — им можно будет делегировать покупку билетов и продуктов, бронирование отелей, планирование поездок и другие бытовые задачи.

Сегодня даже сотрудники без технической подготовки способны создавать простых агентов с помощью интерфейсов-конструкторов, однако для построения надежных, интегрированных и безопасных систем по-прежнему требуется участие инженерных команд.

В перспективе ИИ-агенты могут стать одним из ключевых направлений развития ИИ. На их основе могут появиться корпоративные платформы автоматизации, с помощью которых будет возможно автоматизировать различные бизнес-процессы. Однако этот путь будет долгим, и, вероятно, лишь к 2030 г. мы начнем видеть первые шаги в развитии таких корпоративных платформ, что лишь позднее приведет к их массовому внедрению и зрелости.

Рынок B2B-решений технологий генеративного ИИ

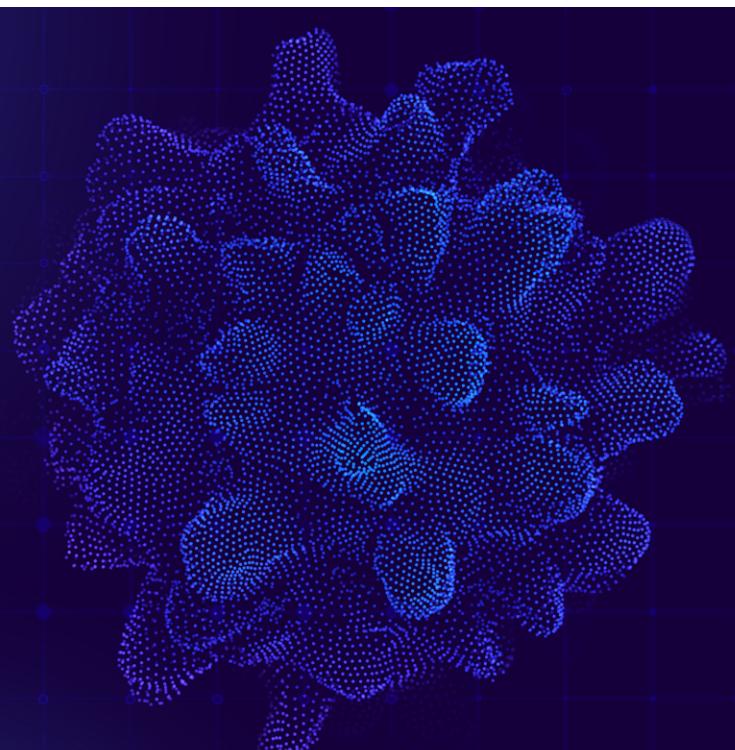
Портрет и структура рынка разработчиков технологий генеративного ИИ

Предложение на рынке генеративного ИИ формируется тремя типами компаний:

- Вендоры генеративного ИИ: компании, специализирующиеся на создании решений на основе генеративных моделей (например, инструменты генерации контента, кода или изображений).
- Разработчики ИИ-помощников: компании, создающие конечные решения на базе генеративных моделей — ассистентов, чат-ботов, RAG-агентов и других систем, выполняющих прикладные задачи и взаимодействующих с внешними источниками данных.
- Мультитехнологические вендоры: компании, сочетающие разработки в области генеративного ИИ с другими ML- и ИИ-направлениями (NLP, CV, рекомендательные системы и т. д.), предлагая широкий спектр решений без узкой специализации.

По результатам опроса вендоров, наибольшую долю рынка занимают разработчики ИИ-помощников — около 34% участников среди опрошенных вендоров по всем технологиям.

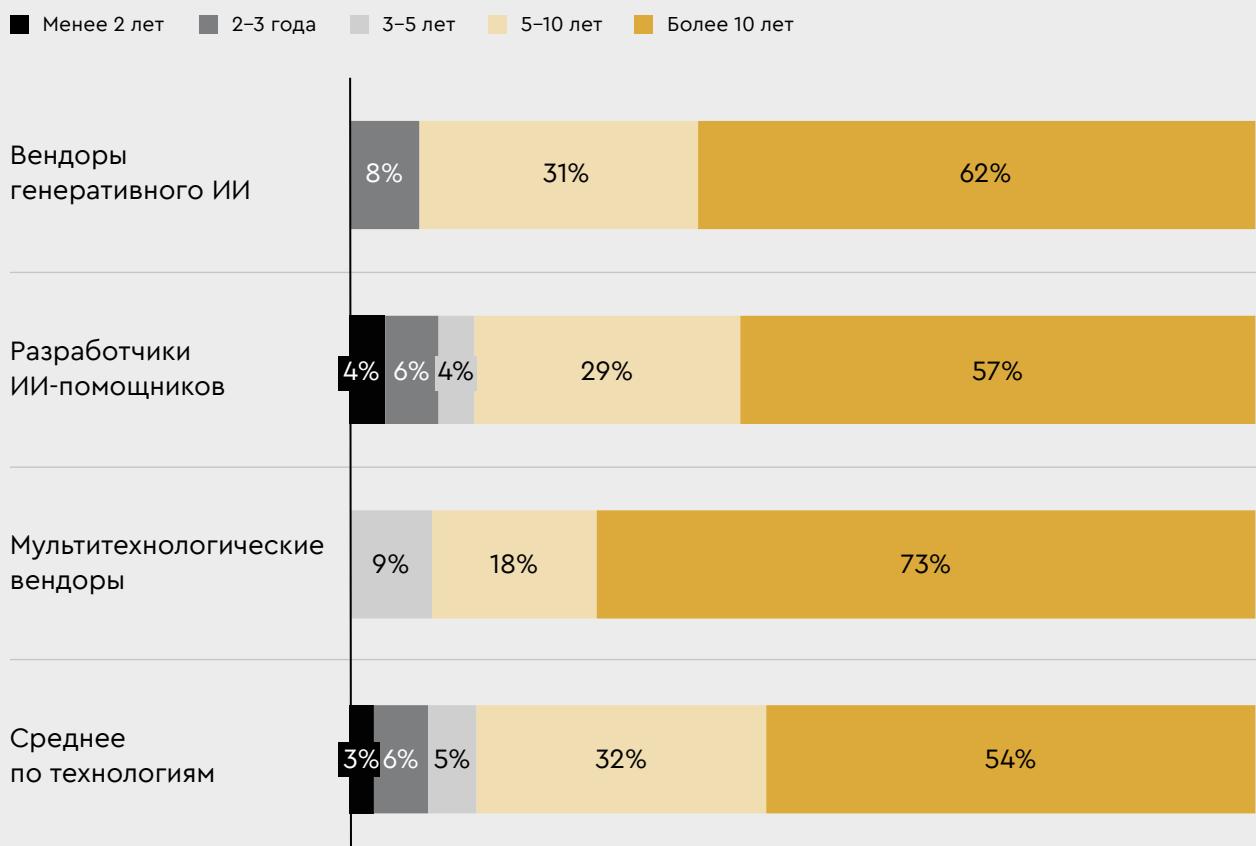
На мультитехнологических вендорах приходится 15%, а на чистых вендоров генеративного ИИ — порядка 9%; остальные 42% приходятся на вендоров речевых технологий и NLP (10%), компьютерного зрения (21%), рекомендательных систем (11%).



Зрелость и опыт вендоров на рынке ИИ-технологий

Вендоры, работающие с технологиями на базе генеративного ИИ, в среднем демонстрируют наиболее высокий уровень зрелости и продолжительности присутствия на рынке. Наибольший опыт на рынке показывают компании с мультитехнологической специализацией (73% работают более 10 лет) и вендоры генеративного ИИ (62%) при среднем уровне по рынку в 54%. Компании, специализирующиеся сегодня на ИИ-помощниках, также сохраняют устойчивое присутствие на рынке — 57% из них работают более 10 лет. При этом большинство из этих игроков пришли в сегмент ИИ-помощников не с нуля: они развивались как ИТ- или ИИ-компании и со временем переориентировали свои компетенции и продуктовые линейки на разработку ИИ-помощников, что обеспечило им сочетание зрелой экспертизы и технологической гибкости.

Время присутствия вендоров на рынке ИИ-технологий, %



Источник: опрос вендоров (блок 1. Контекст, вопрос 1.3. Какое время компания уже присутствует на рынке?)

Размер и структура компаний на рынке генеративного ИИ

Вендоры, специализирующиеся на генеративном ИИ, остаются наиболее компактными: в среднем в таких компаниях работает около 130 сотрудников против 750 в среднем по рынку.

Для сравнения: в компаниях, создающих решения на основе ИИ-помощников, средняя численность достигает 500 человек, а среди мультитехнологических игроков — более 2,5 тыс.

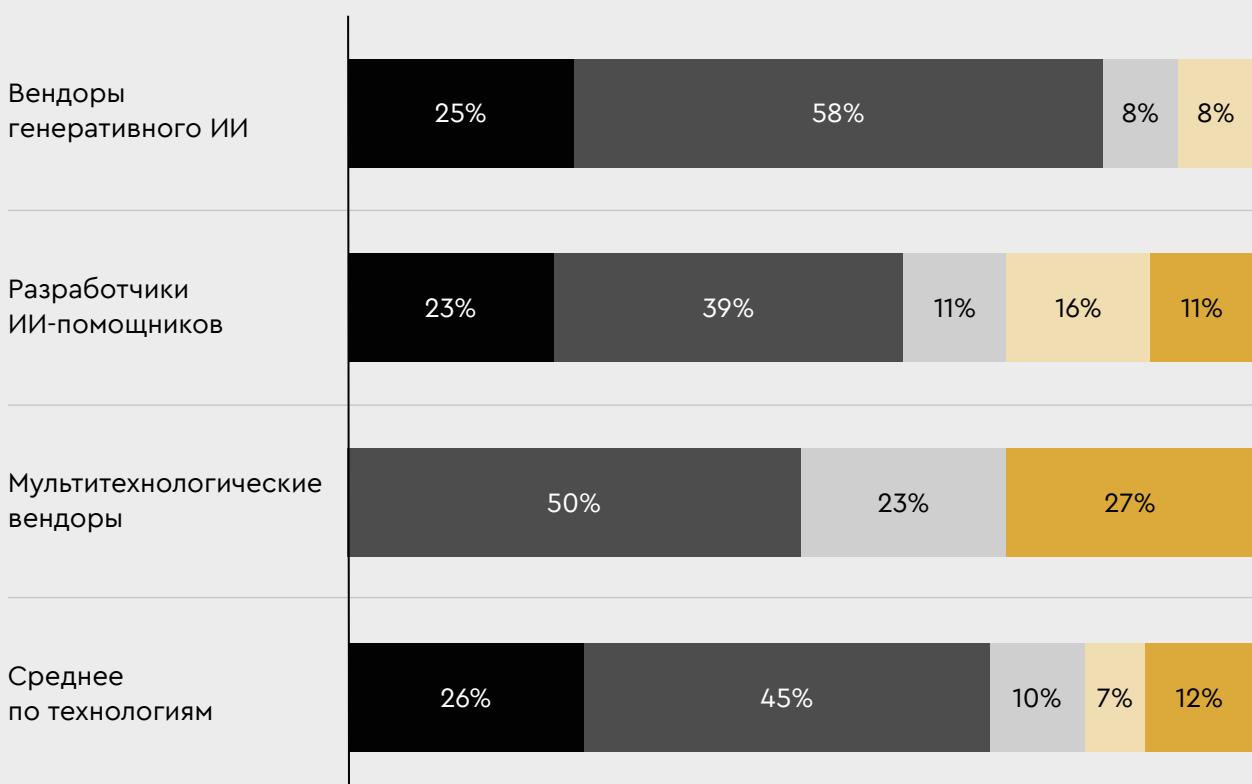
Такое различие связано с уровнем зрелости и стадией развития сегментов. Компании, работающие с генеративным ИИ, часто представляют собой небольшие команды, сфокусированные на разработке конкретных моделей или продуктов.

В то время как мультитехнологические вендоры и разработчики ИИ-помощников — это, как правило, более крупные организации с диверсифицированным портфелем решений и развитой инфраструктурой внедрения.

Несмотря на различия в масштабах, для всех сегментов характерно преобладание компаний среднего размера: наиболее распространенная категория — от 51 до 200 сотрудников. В этот диапазон попадают 58% компаний, специализирующихся на генеративном ИИ, 39% — на ИИ-помощниках и 50% мультитехнологических игроков.

Размер компаний-вендоров по числу сотрудников, %

■ До 50 ■ От 51 до 200 ■ От 201 до 500 ■ От 501 до 1000 ■ Более 1000



Источник: опрос вендоров (блок 1. Контекст, вопрос 1.4. Сколько людей работает у вас в компании?)

Структура выручки в сегментах рынка генеративного ИИ

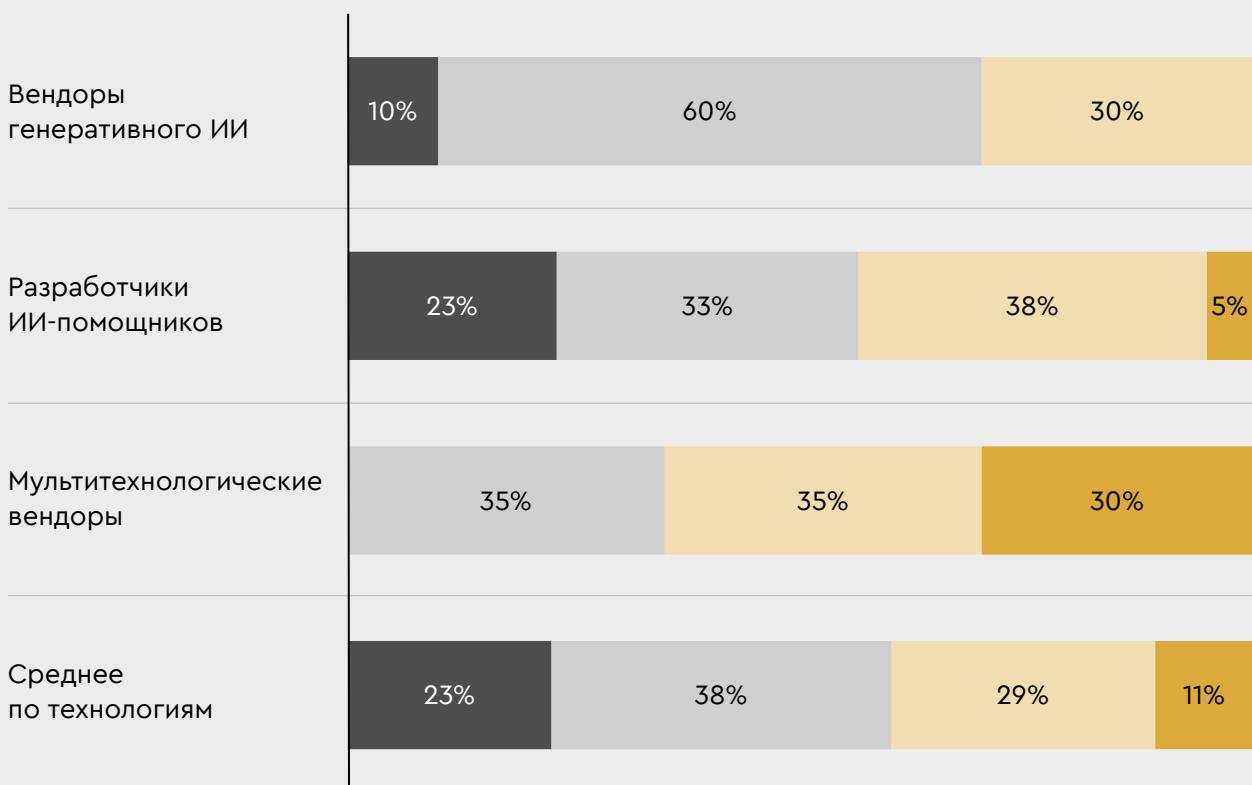
Сегмент ИИ-помощников демонстрирует самые высокие финансовые результаты среди всех направлений: средняя выручка от продажи ИИ-решений у компаний в этой категории достигает почти 1 млрд руб. и превышает показатели даже мультитехнологических вендоров (около 900 млн руб.). Несмотря на то что лишь 5% разработчиков ИИ-помощников имеют выручку выше 1 млрд руб., именно эти компании формируют основную часть объема рынка, аккумулируя значительную долю доходов и поднимая средние показатели до рекордных значений.

Мультитехнологические компании показывают сопоставимые, но более равномерные результаты: около трети из них (30%) имеют выручку от продажи ИИ-решений выше 1 млрд руб., что отражает зрелость и масштаб их бизнеса.

Вендоры, специализирующиеся на генеративном ИИ, напротив, остаются преимущественно компактными по объему выручки: большинство из них (60%) работают в диапазоне 200–500 млн руб. Это связано с тем, что такие компании ориентированы на разработку и совершенствование собственных моделей, а не на крупные интеграционные проекты.

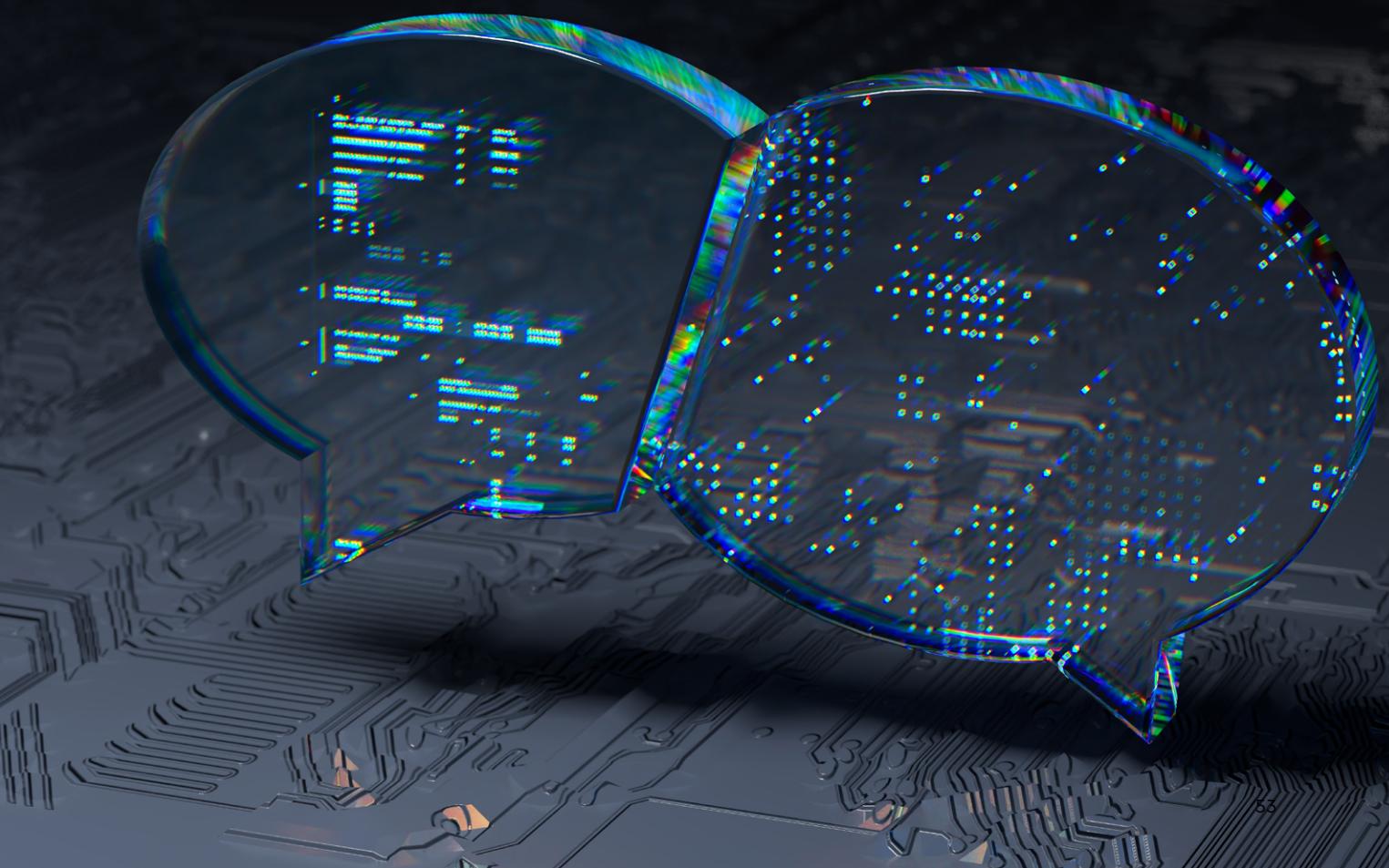
Выручка вендоров по направлениям ИИ, %

■ Менее 200 млн руб. ■ От 200 млн руб. до 500 млн руб. ■ От 500 млн руб. до 1 млрд руб. ■ Более 1 млрд руб.



Источник: опрос вендоров (блок 3. Рыночные позиции, вопрос 3.1. Сколько составляет выручка компании от продажи решений на основе технологии за последний год (2024)?)

Все вендоры, работающие с ИИ-агентами, и 86% компаний, использующих генеративный ИИ, дообучают внешние open-source-модели



Использование и дообучение моделей

Дообучение готовых моделей стало мировым стандартом и является оптимальной практикой для ускоренного внедрения и оптимизации решений. Использование гибридной инициализации, активно применяемое такими компаниями, как Meta, Google, Microsoft, позволяет адаптировать модели под конкретные задачи и сокращать затраты на разработку. По данным опроса вендоров, 14% компаний, работающих с генеративным ИИ, занимаются исключительно дообучением внешних open-source-моделей, в то время как 71% комбинируют использование моделей без изменений с их дообучением. Для России развитие решений на базе open-source-моделей является вполне естественным подходом. Если бы не этот шаг, то на российском рынке значительную долю в использовании заняли иностранные ИИ-сервисы. Лишь 29% разрабатывают собственные базовые модели с нуля, что заметно ниже среднего по другим технологиям (57%). При этом многие компании относят к собственным даже те модели, которые были дообучены на базе внешних весов. Однако степень дообучения может существенно различаться: в одних случаях это лишь адаптация под конкретные задачи, а в других — глубокая переработка, после которой от исходной модели практически ничего не остается.

Похожая тенденция наблюдается и в сегменте ИИ-агентов, где все компании используют дообученные модели и лишь 20% из них имеют собственные разработки. Это объясняется технической сложностью создания агентов и необходимостью интеграции с большим числом внешних систем.

В целом оба сегмента — генеративный ИИ и ИИ-агенты — развиваются преимущественно через адаптацию и интеграцию существующих решений, а не собственную разработку внутри компаний. Такой подход позволяет ускорить вывод продуктов на рынок, снизить издержки на разработку и при этом обеспечить высокий уровень кастомизации под конкретные задачи клиентов.

Внедрение решений генеративного ИИ в организациях

Сегменты потребителей решений на базе генеративного ИИ

По результатам опроса топ-150 вендоров, услуги по внедрению генеративного ИИ в основном ориентированы на компании среднего и крупного масштаба: с ними работают около 85% поставщиков решений. Это объясняется тем, что именно в этих сегментах сосредоточены основные ресурсы и инфраструктура, необходимые для интеграции новых технологий в бизнес-процессы.

Тем не менее на фоне других направлений искусственного интеллекта генеративный ИИ демонстрирует заметно более высокий уровень проникновения в сегмент малого бизнеса. Согласно опросу, 23% вендоров работают с малым бизнесом против 9% в среднем по остальным технологиям. Это делает малый бизнес одним из наиболее перспективных направлений развития рынка генеративного ИИ.

Такое отличие обусловлено самой природой технологии. Генеративный ИИ способен обеспечивать ощутимый эффект даже при ограниченных масштабах бизнеса — ускорять выполнение задач, улучшать клиентский опыт и снижать издержки на коммуникации и аналитику. При этом стоимость внедрения таких решений остается относительно низкой, что делает их доступными для компаний без крупных ИТ-бюджетов. Благодаря этому именно малый бизнес все чаще рассматривает генеративный ИИ как инструмент быстрой цифровой трансформации и повышения операционной эффективности.

Доля вендоров, работающих с компаниями разного масштаба, %



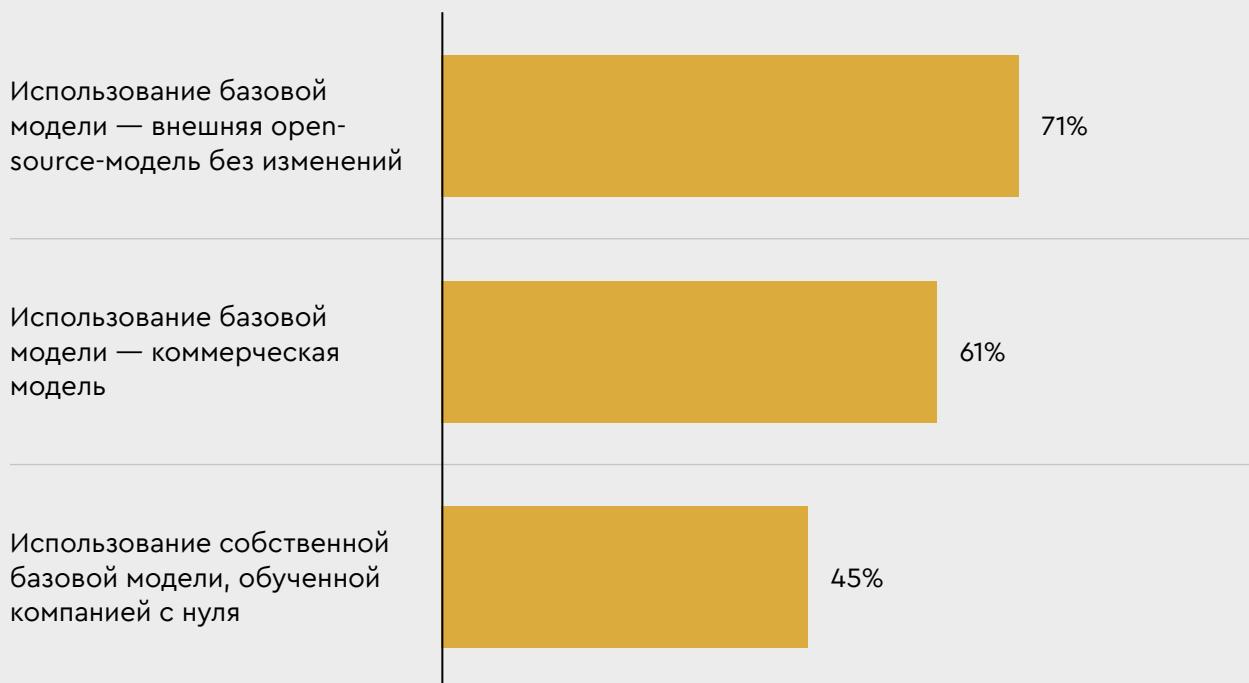
Источник: опрос вендоров (блок 4. Портфель клиентов, вопрос 4.3. Какой тип клиентов преобладает в компании?)

Подходы компаний к внедрению решений на базе генеративного ИИ

Согласно опросу СТО, компании, внедряющие технологии генеративного ИИ, используют два основных подхода. Более 80% организаций делают выбор в пользу специализированных решений, ориентированных на конкретные бизнес-задачи — от маркетинга и клиентского сервиса до аналитики и документооборота. Почти 65% компаний при этом применяют базовые модели, интегрируя их в существующие процессы как основу для собственных разработок или пилотных проектов. Такая структура рынка отражает комбинированную стратегию: специализированные решения обеспечивают быстрый бизнес-эффект и простоту внедрения, тогда как базовые модели требуют большего объема кастомизации, но дают компаниям значительно больше гибкости и возможностей для масштабирования.

При этом на рынке базовых моделей компании используют разные подходы — от внешних open-source-решений до коммерческих продуктов и собственных моделей, обученных с нуля. Наибольшее распространение получили именно open-source-модели без изменений: их применяют более 70% компаний. Это позволяет компаниям ускорять эксперименты с технологией, снижать барьеры для внедрения, быстрее получать прикладной эффект от генеративного ИИ, а также снижать степень зависимости от вендоров.

Типы базовых моделей, используемых компаниями, %



Источник: опрос вендоров (блок 2. Продукт, вопрос 2.4. Какие типы моделей генеративного ИИ компания использует в своих продуктах?)

Экономический потенциал генеративного ИИ в восприятии компаний

Генеративный ИИ, как и другие технологии искусственного интеллекта, в восприятии компаний прежде всего служит инструментом повышения эффективности и роста доходов. Основные ожидания от его внедрения связаны с сокращением прочих операционных затрат и увеличением выручки за счет повышения ценности продукта и улучшения клиентского опыта.

Согласно опросу СТО, 87% компаний, использующих генеративный ИИ, ожидают сокращения прочих видов затрат (в среднем по технологиям — 88%), а 83% — роста выручки за счет улучшения ценности продукта (в среднем — также 83%). При этом только 12% компаний связывают финансовый эффект с сокращением фонда оплаты труда, что заметно ниже не только среднего по технологиям (17%), но и показателя прошлого года — в 2024 г. так считали 49% компаний. Это отражает формирование новой тенденции: генеративный ИИ все чаще рассматривается не как инструмент оптимизации численности персонала, а как средство повышения производительности и перераспределения задач между сотрудниками.

44% компаний связывают влияние генеративного ИИ на выручку с созданием новых продуктов

В то же время генеративный ИИ в большей степени, чем другие технологии, воспринимается как драйвер создания новых продуктов и сервисов. 44% компаний связывают влияние генеративного ИИ на выручку с развитием продуктового портфеля и появлением новых источников дохода — против 29% в среднем по технологиям. Наибольший акцент на этот эффект делают ИТ и технологии (95%), e-commerce (85%) и телеком и медиа (80%).

Генеративный ИИ также является ключевой технологией по ожидаемому экономическому эффекту: компании прогнозируют средний прирост EBITDA на уровне 4% — выше, чем у любых других направлений искусственного интеллекта. Особенно высокие ожидания у e-commerce: совокупный эффект от ИИ, по оценкам, должен достичь 21% EBITDA, из которых около 8% приходится именно на генеративные технологии.

Инвестиции и приоритетность технологии

Генеративный ИИ — лидер по инвестициям: в среднем на него приходится около 4% ИТ-бюджета компаний. В продвинутых в ИИ отраслях этот показатель достигает 5%, тогда как в прочих отраслях — около 3%, но с заметной вариативностью: в отдельных случаях уровень вложений сопоставим с продвинутыми секторами. Это подчеркивает стратегическую значимость генеративного ИИ и широкие возможности его применения в разных индустриях.

Средние инвестиции компаний в технологии ИИ за последний год, в % от годового ИТ-бюджета (сравнение генеративного ИИ со средним значением по прочим технологиям)

 Продвинутые в ИИ отрасли X п. п. Планируемое изменение ИТ-бюджета CV на 2026 г.



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.3. Какой объем средств в % от годового ИТ-бюджета компания уже инвестировала в технологию за последний год? Вопрос 4.5. Какой объем средств в % от ИТ-бюджета компания планирует инвестировать во внедрение технологии на горизонте года?)

Генеративный ИИ остается направлением, демонстрирующим устойчивый рост интереса со стороны бизнеса. Согласно опросу СТО, компании планируют в среднем на 25% увеличить долю инвестиций в генеративный ИИ в следующем году.

Максимальные значения наблюдаются в банковском секторе и ИТ (до 7%), а также в телекоммуникациях и электронной коммерции (около 6%). Это отражает быстрый рост числа прикладных сценариев и универсальность технологии, которая находит применение в самых разных бизнес-моделях.

При этом среди отраслей уже заметны тенденции перераспределения бюджетов в пользу генеративного ИИ. В рейтинге, ИТ и телеком/медиа компании увеличивают долю инвестиций на 1–2 п. п. (до 4–7%), одновременно сокращая вложения в компьютерное зрение (CV) и рекомендательные системы (RecSys) примерно на 1 п. п. Это отражает синергию генеративного ИИ с другими технологиями, такими как CV и RecSys, что позволяет бизнесам оптимизировать ресурсы и получать максимальный эффект от всех технологий в совокупности.



Будущее и перспективы развития генеративного ИИ

Будущее развитие технологии

80% вендоров и 65% индустриальных компаний ожидают существенного расширения возможностей генеративного ИИ и появления новых решений на его базе в ближайшие годы

По результатам опроса вендоров ИИ-технологий, 80% участников рынка прогнозируют значительный рост возможностей и появление новых решений на базе генеративного ИИ в ближайшие годы. Аналогичные ожидания демонстрируют и компании: согласно опросу СТО, 65% организаций ожидают дальнейшего развития технологии, а еще 10% относят ее к стадии зарождения. Таким образом, и бизнес, и разработчики сходятся во мнении, что потенциал генеративного ИИ еще далеко не исчерпан.

При этом оценки темпов развития во многом совпадают. Резкий рост применения генеративного ИИ ожидают 45% вендоров, а 15% считают, что технология со временем станет ключевой во всех бизнес-процессах. Аналогичные ожидания демонстрируют и компании: 70% прогнозируют увеличение масштабов использования генеративного ИИ (в том числе 30% — резкий рост), а 10% уверены, что в будущем генеративный ИИ станет основой корпоративных процессов.

Оптимизм обеих сторон отражается и в более высоких рыночных прогнозах. Вендоры, специализирующиеся на генеративном ИИ, оценивают средние темпы роста рынка ИИ-технологий в целом выше, чем представители других направлений, — около 30% в год против 25%. При этом 10% из них оценивают темпы роста рынка ИИ-технологий на уровне свыше 40%.

Таким образом, обе стороны рынка — как индустриальные компании, так и разработчики — едины в оценке будущего технологии. Развитие генеративного ИИ только набирает темп и, по ожиданиям участников рынка, станет главным катализатором следующего этапа роста корпоративных ИИ-решений в России.

Результаты опроса населения России также подтверждают эти прогнозы. Подавляющее большинство (79%) ожидает роста использования генеративного ИИ. Каждый четвертый (26%) из этой группы (среди ИТ-специалистов — 27%) уверен, что технология станет ключевой, используемой во всех бизнес-процессах и заменит значительную долю других решений.

Мы предполагаем, что к 2030 г. генеративный ИИ будет активно использоваться более чем половиной населения планеты. Рост применения технологии будет происходить как вширь, с увеличением числа пользователей, так и вглубь, с проникновением ИИ в повседневную жизнь и созданием новых бизнес-кейсов. Это приведет к тому, что генеративный ИИ станет неотъемлемой частью различных сфер — от ИИ-ассистентов до более сложных решений в профессиональной деятельности и личной жизни.

Барьеры развития технологии

Несмотря на высокий оптимизм и ожидания в отношении дальнейшего роста генеративного ИИ, компании и вендоры отмечают, что развитие технологии сталкивается с рядом системных ограничений. Эти барьеры различаются по своей природе, но в совокупности формируют ключевые вызовы для масштабирования решений на базе генеративного ИИ в России.

Со стороны вендоров основными препятствиями в развитии генеративного ИИ выступают высокая стоимость поддержания технологического лидерства — на это указали 91% участников опроса вендоров ИИ-технологий, а также поиск и удержание квалифицированных технических специалистов (56%). Дефицит экспертизы в ИИ остается одной из ключевых проблем во всем мире: спрос на специалистов, способных проектировать и обучать сложные модели, кратно превышает предложение, что усиливает конкуренцию за кадры. В России к кадровому дефициту добавляются инфраструктурные факторы: ограниченный доступ к «железу» и сложности с его ввозом повышают стоимость экспериментов и удлиняют цикл разработки. В результате компании чаще выбирают более консервативный курс — делают ставку на решения с доказанной пользой, а не на рискованные R&D-проекты. Это не закрывает путь к технологическому лидерству, но делает его более ресурсоемким: сложнее экспериментировать и системно финансировать высокорисковые инициативы. Критично не только нанимать сильных специалистов, но и создавать условия для их роста: доступ к вычислительным ресурсам, исследовательские «песочницы», гибкие треки развития и внутренняя подготовка ML-инженеров. Это ускоряет внедрение и улучшение технологий несмотря на внешние ограничения.

Со стороны компаний ключевым ограничением внедрения является высокая финансовая стоимость. Наибольшая доля респондентов — 75% компаний, участвовавших в опросе СТО, — указывает именно этот фактор как основной барьер масштабирования технологий. Если в 2024 г. лишь 12% респондентов отмечали финансовую сложность как значимое препятствие, то в 2025 г. этот барьер вышел на первый план. Это свидетельствует о том, что по мере перехода компаний от пилотов к более масштабным внедрениям растет потребность в инвестициях в инфраструктуру, вычислительные мощности и адаптацию бизнес-процессов.

Тем не менее уклон в сторону использования Small LM (малых моделей для специфических задач), а также использования генеративных моделей на всех этапах — от архитектуры до инференса — может помочь снизить эту нагрузку.

Кроме того, с развитием облачных решений стоимость использования ИИ значительно снижается: так, например, цена за токен с GPT-4 (начало 2023 г.) до GPT-4o (середина 2024 г.) упала примерно в 150 раз, что также способствует сокращению затрат на внедрение.

Высокая стоимость становится новым вызовом для рынка, который предстоит преодолеть, чтобы обеспечить устойчивое развитие экосистемы генеративного ИИ в России.

Отдельной группой барьеров выступают институциональные ограничения в критически важных отраслях, таких как банкинг и страхование, медицина и здравоохранение, а также другие ключевые сектора. Процесс внедрения новых технологий здесь остается длительным и сложным вследствие высокой степени бюрократизации и многоуровневых согласований, требуемых на каждом этапе — от пилотных проектов до масштабирования. В отличие от менее регламентированных отраслей, где решения могут масштабироваться оперативно, в таких сферах внедрение ИИ затруднено отсутствием единой технологической платформы, что замедляет реализацию проектов.

Еще одним возможным барьером для развития генеративного ИИ является стремление людей наделить технологии человеческими чертами. Многие воспринимают ИИ как интеллект, который должен имитировать человеческие когнитивные способности. Однако более оптимальным и эффективным подходом может стать развитие генеративного ИИ в направлении, не предполагающем имитацию человеческих характеристик. Модели, ориентированные на выполнение конкретных задач без попытки воспроизвести человеческое поведение, могут обеспечить более точное и быстрое решение проблем, что повысит их эффективность и упростит интеграцию в различные бизнес-процессы.

Рекомендации для ускоренного преодоления барьеров

Одним из возможных вариантов преодоления барьеров высокой стоимости разработки технологий может стать переход к платформенной модели. Компании вместо создания собственных моделей могут сосредоточиться на использовании крупных LLM-платформ и разработке прикладных SaaS-решений на их основе. Это позволит сократить издержки, ускорить внедрение технологий и сформировать устойчивую экосистему, где несколько ведущих вендоров станут технологическим ядром рынка.

В целях преодоления выявленных барьеров государству целесообразно стимулировать подготовку и реализацию пилотных проектов по внедрению ИИ в приоритетных сферах. Одновременно необходимо расширить меры экономической поддержки отраслей и компаний — гранты и субсидии на НИОКР и внедрение, налоговые стимулы, льготное финансирование — для снижения совокупных издержек разработки и опытной эксплуатации решений и ускорения вывода технологий в новые кейсы.

Регуляторная политика должна строиться на принципах рискоориентированного и технологически нейтрального регулирования: соразмерные требования к безопасности и защите данных, использование пилотных правовых режимов — при недопущении необоснованного ограничения инновационной активности.

Таким образом, несмотря на сохраняющиеся финансовые, кадровые и институциональные ограничения, уверенность бизнеса и вендоров в потенциале генеративного ИИ, а также высокий уровень инвестиций и интереса к технологии позволяют ожидать, что рынок сможет успешно преодолеть существующие барьеры и ускорить масштабное внедрение решений.

Обработка естественного языка (NLP) и речевые технологии

Развитие NLP и речевых технологий

Обработка естественного языка и речевые технологии — ключевые направления искусственного интеллекта, направленные на взаимодействие человека и устройств с использованием естественной речи. NLP отвечает за анализ и генерацию текста, а речевые технологии — за распознавание, синтез и обработку звуковой речи.

Развиваясь с 1950-х гг., эти технологии эволюционировали от раздельных решений к единой системе, объединяющей обработку текста и речи.

1950–1969 гг.

Первый этап: обработка текста на основе правил

Первые эксперименты в области обработки языка были сосредоточены на машинном переводе, для которого ранние системы использовали правила и словари, созданные лингвистами. В 1954 г. компьютер IBM 701 продемонстрировал первый автоматический перевод с русского на английский.

1970–1989 гг.

Переход к вероятностному моделированию и первые системы распознавания речи

К 1970-м гг. использование только правил в NLP стало недостаточным для решения сложных задач. Исследователи начали учитывать семантику и контекст, переходя к вероятностным моделям. В 1980-х гг. команда IBM Research разработала IBM Models, что стало основой статистического машинного перевода (SMT).

1970-е гг. также стали важным десятилетием для автоматического распознавания речи: система HARPY впервые применила алгоритм beam search для эффективного разбора вариантов. Позднее вехой развития стала Sphinx (1987) — первая система непрерывного распознавания речи, не зависящая от диктора.

1990–2009 гг.

Этап больших данных и массовых продуктов

В 1990–2000-е гг. машинное обучение стало основным инструментом в NLP и речевых технологиях, а рост интернет-данных и развитие лингвистических ресурсов способствовали переходу от экспериментальных систем к прикладным решениям.

На рубеже 2000-х и 2010-х гг. технологии обработки языка стали доступны широкой аудитории. Диалоговые системы и голосовые ассистенты эволюционировали от IVR-систем к современным помощникам, а в 2006 и 2011 гг. были запущены Google Translate и Яндекс Переводчик, сделавшие автоматический перевод массовым сервисом.

2010–2019 гг.

Глубокое обучение и слияние NLP и речевых технологий

С начала 2010-х гг. развитие обработки языка и речи ускорилось благодаря глубоким нейронным сетям. Архитектура Seq2Seq с механизмом внимания (2014) дала прорыв в машинном переводе, а трансформеры (2017) стали основой для современных языковых моделей.

Качество распознавания речи значительно улучшилось: в 2016 г. системы достигли паритета с человеком, а границы между NLP и речью начали стираться — голосовые ассистенты стали работать в связке: распознавание речи → языковая модель → генерация ответа → TTS (Text-to-Speech).

2020–2025 гг.

Замещение классических подходов генеративными моделями

С распространением генеративного ИИ произошло преобразование NLP и речевых технологий. Генеративные модели начали заменять классические подходы, выполняя задачи анализа, перевода, распознавания и синтеза речи в единой архитектуре. Одновременно развивались открытые инициативы, такие как Common Voice от Mozilla, собравшие десятки тысяч часов речи на более чем 100 языках, что улучшило качество моделей и расширило языковой охват, включая редкие языки.

Ключевые тренды и изменения в NLP и речевых технологиях

Современное развитие NLP и речевых технологий определяется переходом к более мощным моделям и растущей ролью генеративного ИИ, расширяющего возможности анализа, синтеза и мультимодального восприятия. В 2025 г. можно выделить несколько ключевых трендов, определяющих дальнейшее развитие NLP и речевых технологий.

- **Увеличение роли генеративных моделей.** Крупные языковые модели (LLM) все чаще заменяют традиционные NLP-решения, такие как BERT (Bidirectional Encoder Representations from Transformers), предлагая более гибкий и контекстно зависимый подход к обработке текста.
- **Рост мультимодальности.** NLP все чаще интегрируется в мультимодальные архитектуры, объединяющие обработку текста, звука и изображений в единой модели. Такие решения позволяют системам лучше понимать контекст и обеспечивать более естественное взаимодействие с пользователем. Однако внедрение мультимодальных моделей остается капиталоемким из-за высокой стоимости вычислений и сложности обучения.
- **Голосовая экспрессия и эмоциональная вариативность.** Современные речевые модели начинают осваивать способность воспроизводить речь с различными интонациями и настроениями, даже если эти состояния не были явно заданы в обучающем наборе. Компании активно преодолевают барьер ограниченной способности моделей адаптировать эмоциональную окраску речи под контекст и намерение пользователя, что расширяет применение технологий синтеза речи и повышает естественность взаимодействия человека с ИИ. При этом бизнес целенаправленно движется к эмоционально адаптивной речи, но реализуют это пока не все — чаще в отдельных продуктах и пилотах. Как устойчивый тренд направление сохраняется: спрос на вариативность голоса растет, а кейсы постепенно выходят за рамки экспериментов.

Основные области внедрения речевых технологий и NLP

Технологии обработки естественного языка и речи сегодня зачастую применяются в связке с другими направлениями искусственного интеллекта — генеративными моделями и рекомендательными системами, что позволяет создавать более комплексные интеллектуальные решения. Ключевые функции технологий включают распознавание и синтез речи, анализ и понимание текста, машинный перевод, извлечение информации, распознавание эмоций и голоса, а также создание диалоговых и голосовых интерфейсов, обеспечивая естественное взаимодействие человека с компьютером.

По результатам проведенного опроса, наиболее широкое применение речевые технологии и NLP находят в клиентском сервисе (68% компаний). Это объясняется тем, что коммуникация с клиентами является одной из ключевых точек взаимодействия бизнеса с потребителем, но при этом является достаточно трудозатратной и требует автоматизации для повышения эффективности. При этом различия между продвинутыми в ИИ отраслями и остальными незначительны (72% против 66%), что связано со зрелостью и длительной историей развития технологии: большинство компаний уже интегрировали ее в базовые процессы. В то же время в наиболее технологически развитых секторах генеративные модели постепенно проникают в те функции, которые ранее решались с использованием классических NLP и речевых технологий, обеспечивая более гибкое и контекстуальное взаимодействие с пользователями.

Зрелость технологии также проявляется в ее узнаваемости среди обычных пользователей. 53% опрошенных знакомы с технологиями распознавания и синтеза речи, однако только немногие из них хорошо разбираются в предмете. Такая высокая узнаваемость объясняется ежедневным взаимодействием с технологией, например через использование голосовых помощников, таких как Алиса или Siri, и сервисов озвучивания текста, а также более чем десятилетним присутствием технологий на рынке, что подтверждается тем, что 75% компаний внедрили эти решения более года назад.

Помимо клиентского сервиса, речевые технологии и NLP активно применяются на производстве, например для автоматизации мониторинга процессов, анализа технической документации и управления качеством. 31% компаний в продвинутых в ИИ отраслях и 45% в прочих уже внедрили решения на их основе. Значительно более ощутимы различия в маркетинге и продажах: 73% компаний из продвинутых отраслей используют технологии против 30% в остальных. Это связано с тем, что в данной функции речевые технологии и NLP чаще всего применяются в бизнес-кейсе виртуальных ассистентов, которые сложны в разработке и требуют интеграции с внутренними системами и данными.

Доля компаний, внедривших речевые технологии и NLP в бизнес-функции, от общего числа опрошенных компаний по индустрии, %

	Клиент- ский сервис	Произ- водство	Марке- тинг и прода- жи	HR	ИТ	Исследо- вания и разра- ботки	Внутрен- ние комму- никации	Стратегия	Финансы	Цепочки поставок	Среднее кол-во функций
ИТ и технологии		100%	83%	100%	75%	75%					4,3
Химия и нефтехимия		67%	67%		100%	100%					3,3
Электроэнергетика		100%	100%					100%			3,0
Металлы и горная добыча		84%	60%	50%	67%						2,6
Банкинг и страхование		100%		83%	40%	25%					2,5
E-commerce		87%		67%			50%				2,0
Медицина и здравоохранение		100%	100%								2,0
Ритейл		67%	29%	57%		17%		17%			1,9
Автомобильная промышленность		78%	100%								1,8
Сельское хозяйство		75%		100%							1,8
Машиностроение		33%	29%				100%				1,6
Строительство и недвижимость		92%		50%							1,4
Транспорт и логистика		100%	33%								1,3
FMCG				100%							1,0
Телеком и медиа			40%	40%							0,8
Нефть и газ			25%								0,3
Продвинутые в ИИ		72%	31%	73%	29%	25%	13%	0%	0%	0%	2,4
Обычные отрасли		66%	45%	30%	14%	10%	8%	10%	0%	0%	1,8
В среднем		68%	42%	40%	18%	14%	9%	7%	0%	0%	2,0

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологий, вопрос 2.1. В каких функциях внедряется данная технология и в какой степени?)

По результатам опроса были выделены топ-5 наиболее распространенных кейсов применения речевых технологий и NLP:

Топ-5 бизнес-кейсов использования речевых технологий и NLP (1/2)

Топ-1 бизнес-кейс по популярности в использовании

Генеративный ИИ, речевые технологии и NLP

Цифровой виртуальный ассистент для взаимодействия с клиентами (маркетинг и продажи)

Описание

ИИ-ассистент сопровождает клиента на всем пути покупки — консультирует по продуктам, выполняет функции перекрестных продаж. Система обучается на данных о поведении клиентов и CRM, обеспечивая персонализированные рекомендации и повышение конверсии продаж

Результат

Повышение качества обслуживания и рост удовлетворенности клиентов

Частота использования*, %



Топ-2 бизнес-кейс по популярности в использовании

Генеративный ИИ, речевые технологии и NLP, RecSys

Автоматизированный чат-бот клиентской поддержки (клиентский сервис)

Описание

Чат-бот на основе ИИ принимает обращения клиентов, отвечает на типовые запросы, помогает решать распространенные проблемы. Система интегрируется с базами знаний, CRM и внутренними системами компании, обеспечивая быструю и точную обработку запросов без участия оператора

Результат

Сокращение нагрузки на поддержку и повышение скорости обслуживания

Частота использования*, %



Топ-3 бизнес-кейс по популярности в использовании

Генеративный ИИ, речевые технологии и NLP

Поиск товаров по фото (маркетинг и продажи)

Описание

Система с помощью ИИ анализирует тексты отзывов и обращений, выявляет темы, эмоции и причины недовольства. Отчеты формируются автоматически и помогают выявлять проблемные зоны и точки роста сервиса

Результат

Повышение качества обслуживания и рост удовлетворенности клиентов

Частота использования*, %



* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

Топ-5 бизнес-кейсов использования речевых технологий и NLP (2/2)

Топ-4 бизнес-кейс по популярности в использовании

Генеративный ИИ,
речевые технологии и NLP

Автоматическая транскрибация и суммаризация встреч (внутренние коммуникации)

Описание

Система автоматически расшифровывает аудио- и видеозаписи встреч, выделяет ключевые темы, решения и задачи. Итоги формируются в виде краткого отчета и могут автоматически направляться участникам

Результат

Экономия времени и повышение прозрачности внутренних коммуникаций

Частота использования*, %



Топ-5 бизнес-кейс по популярности в использовании

Генеративный ИИ,
речевые технологии и NLP

Анализ качества работы колл-центра: выявление повторяющихся проблем, оценка взаимодействия (клиентский сервис)

Описание

ИИ анализирует записи разговоров и чаты, выявляя повторяющиеся проблемы, эмоциональный фон и качество взаимодействия операторов с клиентами. Система формирует отчеты и рекомендации для улучшения сервиса и обучения сотрудников

Результат

Повышение качества обслуживания и снижение числа повторных обращений

Частота использования*, %



* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

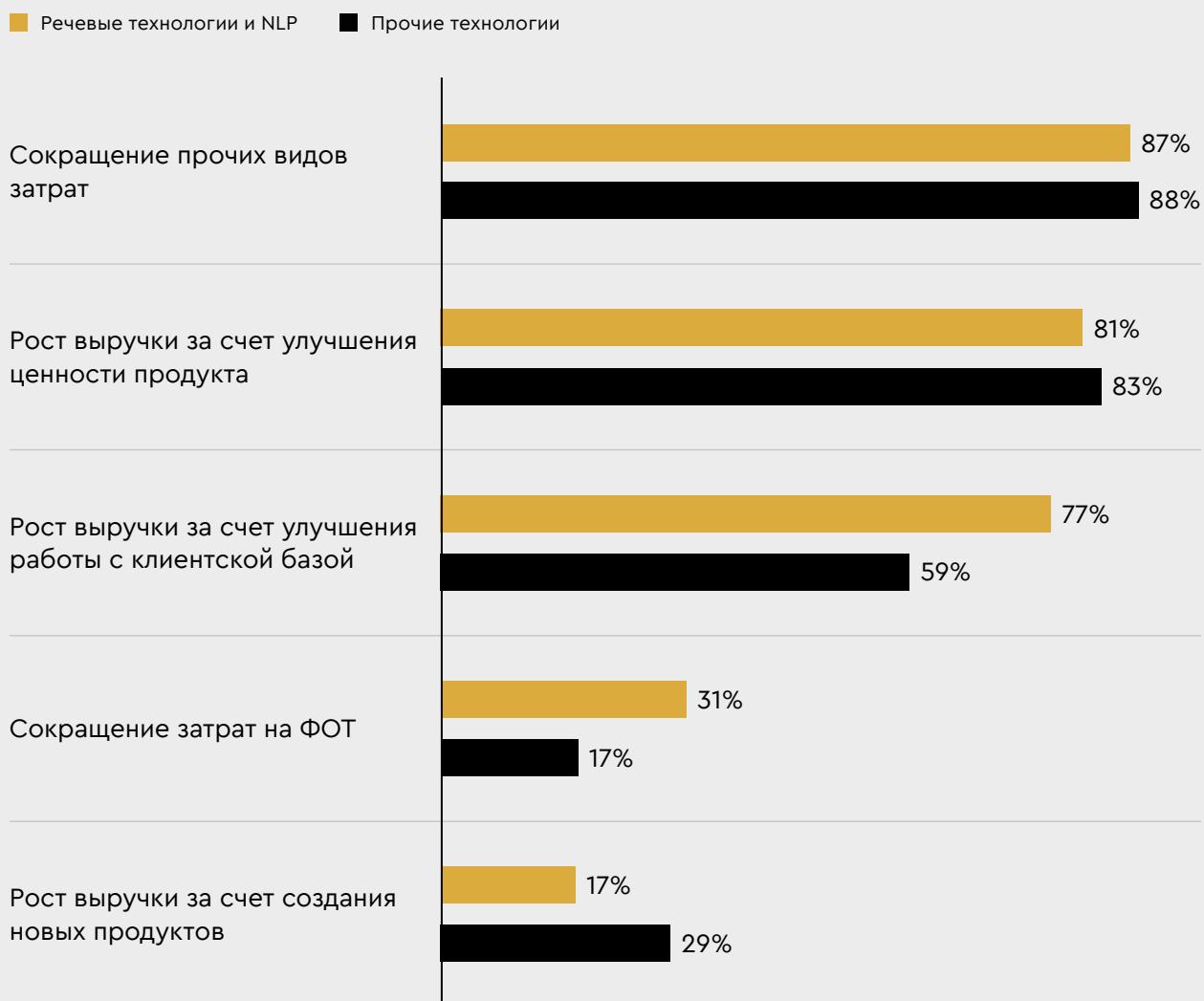
Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

Применение речевых технологий и NLP в этих ключевых бизнес-кейсах также активно развивается с использованием генеративного ИИ. Однако основным фактором выбора остается баланс между качеством и стоимостью решения. В то время как генеративные модели предлагают более гибкие и мощные возможности, традиционные NLP-решения все еще остаются более кост-эффективными для ряда задач, где высокая точность и контекстность не являются критически важными.

Внедрение решений речевых технологий и NLP в организациях

Для компаний, использующих речевые технологии и NLP, особенно значим эффект от работы с клиентской базой: 80% компаний отмечают улучшение качества и скорости взаимодействия с клиентами. Кроме того, эти решения чаще других технологий связывают с сокращением фонда оплаты труда (30%), что обусловлено автоматизацией процессов в колл-центрах и службах поддержки. Использование голосовых и текстовых ассистентов позволяет перераспределять нагрузку между операторами, снижая операционные издержки и одновременно повышая качество обслуживания.

Сложности, которые видят компании в использовании технологий, % ответов (сравнение CV со средним значением по прочим технологиям)



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.1. В чем компания видит источник финансового эффекта от внедрения технологий?)

NLP воспринимается бизнесом как технология с высоким экономическим потенциалом: 90% компаний из продвинутых в ИИ отраслей отмечают ее способность обеспечивать значимый финансовый эффект в отдельных функциях. Особенно высоко ценность NLP оценивается в ИТ и e-commerce, где все компании указывают на ее прямое влияние на эффективность и выручку — прежде всего за счет автоматизации коммуникаций, интеллектуального поиска и персонализации клиентского опыта.

В количественном выражении продвинутые отрасли ожидают прироста EBITDA на уровне до 4% в 2025 г., тогда как в прочих секторах прогнозируется более умеренный эффект — около 2%, что отражает различие в масштабе внедрения и зрелости применения технологии.

Средний ожидаемый эффект компаний от внедрения речевых технологий и NLP на 2026 г., % EBITDA

▣ Продвинутые в ИИ отрасли

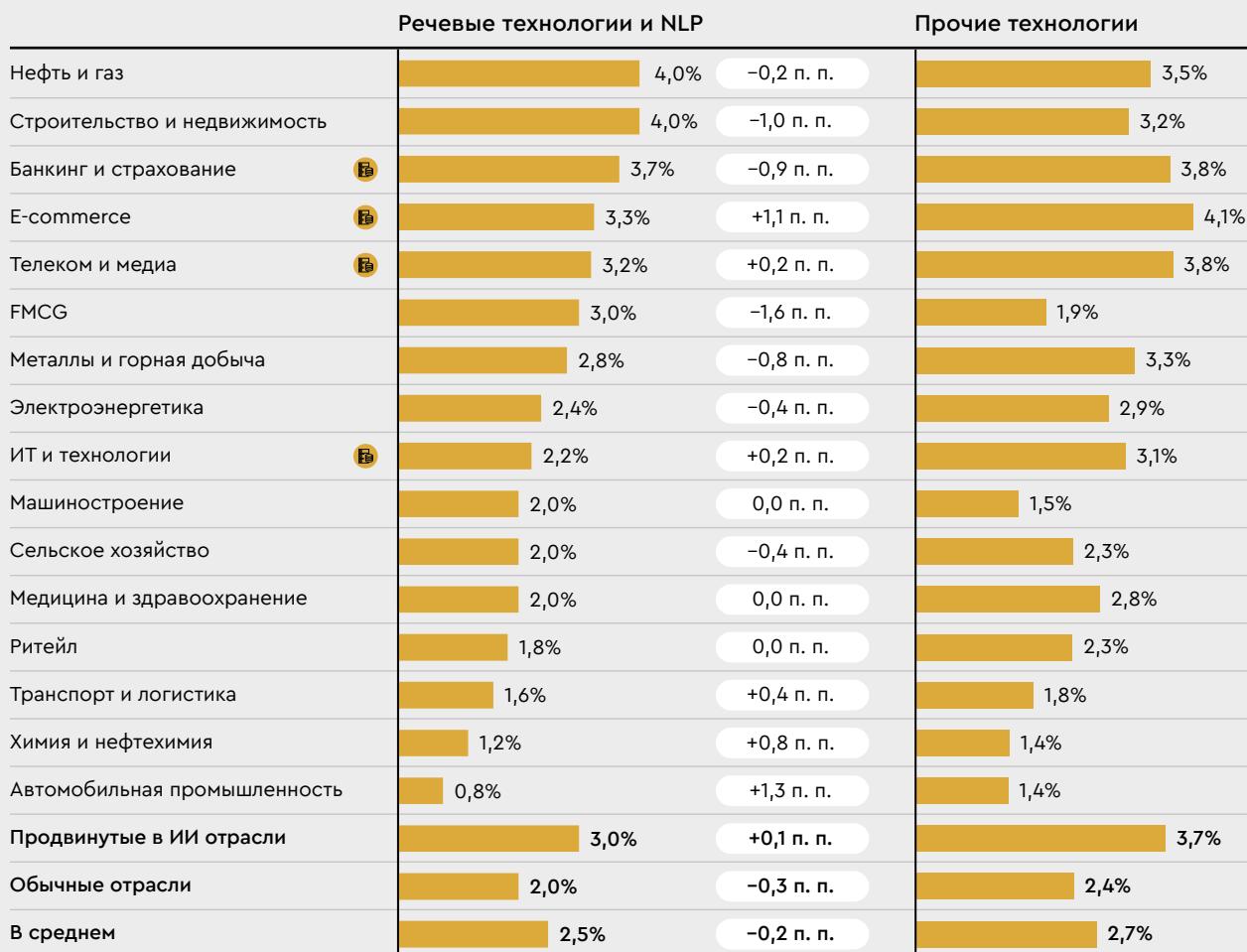


Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.2. Какой финансовый эффект компания уже получила с момента внедрения технологии в % от годовой EBITDA за последний год?)

По результатам анализа инвестиционной активности, осознание экономического потенциала NLP-технологий отражается в планах компаний на следующий год. Вложения сохраняются на стабильном уровне — около 3% ИТ-бюджета в продвинутых отраслях и 2% в прочих. При этом ряд индустрий демонстрируют тенденцию к наращиванию инвестиций: в e-commerce доля затрат на NLP достигнет 4%, увеличившись на 1 п. п., а в автомобильной промышленности и химии/нефтехимии — до 2%, увеличившись также на 1 п. п. Дополнительным фактором такого роста, помимо увеличения бизнес-кейсов применения технологии, может выступать интеграция генеративного ИИ в существующие речевые технологии и NLP-решения, что расширяет функциональность технологий и повышает их прикладную ценность для бизнеса.

Средние инвестиции компаний в технологии ИИ за последний год, в % от годового ИТ-бюджета (сравнение речевых технологий и NLP со средним значением по прочим технологиям)

Продвинутые в ИИ отрасли Х п. п. Планируемое изменение ИТ-бюджета СV на 2026 г.



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.3. Какой объем средств в % от годового ИТ-бюджета компания уже инвестировала в технологию за последний год? Вопрос 4.5. Какой объем средств в % от ИТ-бюджета компания планирует инвестировать во внедрение технологии на горизонте года?)

Рынок B2B-решений речевых технологий и NLP

Проведенный опрос показал, что среди 150 российских вендоров, принявших участие в исследовании, 10% специализируются на речевых технологиях и NLP.

Рынок вендоров речевых технологий и NLP характеризуется высокой дифференциацией по объему выручки.

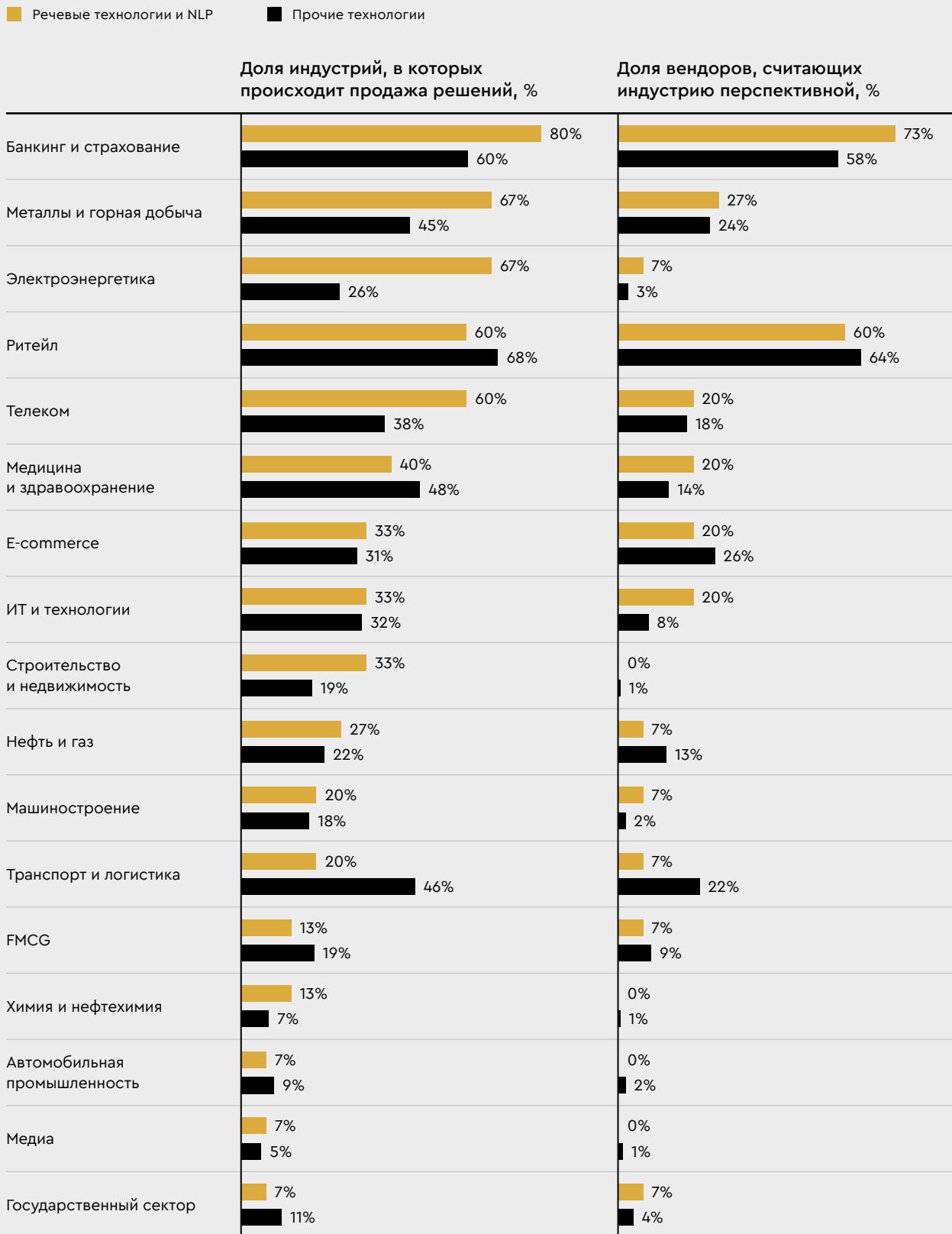
На нем присутствуют и малые компании с выручкой менее 200 млн руб. (18%), и средние — от 200 до 500 млн руб. (45%), и крупные — с оборотом свыше 1 млрд руб. (около 10%). Средняя выручка по рынку составляет порядка 570 млн руб., что выше, чем у вендоров генеративного ИИ, компьютерного зрения и рекомендательных систем (в среднем около 350 млн руб.). Это указывает на устойчивый спрос и зрелость сегмента NLP и речевых технологий.

С точки зрения спроса на услуги структура клиентов в сегменте речевых технологий и NLP заметно отличается от других направлений искусственного интеллекта. Вендоры, специализирующиеся на речевых решениях, в большинстве случаев работают с компаниями банковского и страхового сектора (80%), где технологии применяются для автоматизации колл-центров и поддержки клиентского сервиса. В отличие от других ИИ-направлений, существенная доля заказчиков также приходится на отрасли металлов и горной добычи, а также электроэнергетику, что может быть связано с необходимостью обеспечения безопасности, мониторинга процессов и автоматизации внутренней коммуникации на производственных площадках.

Вендоры NLP и речевых технологий считают наиболее перспективными отраслями те, где активно происходит взаимодействие с клиентами, такие как банкинг и страхование (73%) и ритейл (60%). В банкинге и страховании потенциал обусловлен использованием голосовых ассистентов для улучшения взаимодействия с аудиторией и повышения качества клиентского сервиса. В ритейле технологии NLP и речевые технологии направлены на автоматизацию обслуживания клиентов и улучшение покупательского опыта, что способствует повышению эффективности бизнес-процессов и удовлетворенности потребителей.

Вендоры, фокусирующиеся на NLP, преимущественно сотрудничают со средним и крупным бизнесом: среди их клиентов отсутствуют компании малого сегмента. Для малого бизнеса внедрение речевых и NLP-технологий остается неприоритетным направлением, поскольку объем клиентских коммуникаций сравнительно невелик и не требует привлечения внешних решений.

Индустрии, в которых происходят продажи у вендоров/доля вендоров, считающих индустрию перспективной, % ответов (сравнение речевых технологий и NLP со средним значением по прочим технологиям)



Источник: опрос вендоров (блок 4. Портфель клиентов, вопрос 4.1. В каких индустриях компания продает решения на основе технологий и какие считает наиболее перспективными?)

Будущее и перспективы развития речевых технологий и NLP

Речевые технологии и NLP воспринимаются компаниями как зрелые, но еще не исчерпавшие свой потенциал: 60% респондентов ожидают значительного развития и появления новых решений в ближайшие годы, тогда как 40% считают, что основные возможности уже реализованы. Среди вендоров распределение ожиданий обратное: 40% прогнозируют дальнейшее развитие, а 60% полагают, что технология достигла пика. Тем не менее обе стороны сходятся во мнении, что применение речевых технологий и NLP будет продолжать постепенно расти. Это подтверждается оценками 46% компаний и 57% вендоров, ожидающих плавного увеличения использования технологий за счет расширения отраслевого проникновения.

Эти ожидания также находят отражение в опросе населения России: 80% респондентов считают, что значимая часть потенциала технологии реализована, но решения еще будут активно улучшаться. Более того, 14% в целом и четверть среди ИТ-специалистов считают, что технология станет ключевой, используемой во всех бизнес-процессах и замещающей значительную долю других решений.

Барьерное поле можно условно разделить на две группы — ограничения, характерные для классических NLP-решений, и барьеры, с которыми сталкиваются речевые технологии.

Для NLP основными препятствиями развития остаются:

- ограниченность прикладных сценариев, где технология может быть использована (чаще точечные задачи, а не сквозные процессы);
- высокая длительность развертывания и интеграции решений в корпоративные процессы, что усложняет масштабирование.

Для речевых технологий дальнейшее развитие сегодня во многом определяется возможностью интеграции генеративного ИИ в голосовые сценарии. Именно барьеры для внедрения генеративного ИИ становятся ключевым фактором, сдерживающим рост применения голосовых решений в компаниях:

- повышенная задержка отклика при работе крупных моделей (2–3 секунды), которая в сочетании с распознаванием и синтезом речи формирует заметную паузу в диалоге;
- ограниченная возможность контроля качества и предсказуемости ответа в голосовом потоке, поскольку дополнительные механизмы валидации приводят к увеличению задержки

При этом решения на фоне этих ограничений уже начинают появляться. Например, Yandex Cloud в 2025 г. представил инструмент для голосовых сценариев, который формирует ответ менее чем за одну секунду, что значительно повышает плавность и естественность диалога. Дополнительно компании все чаще используют узкоспециализированные голосовые сценарии, где контекст и набор возможных ответов заранее ограничены, что позволяет уменьшить риск некорректных формулировок без увеличения задержки.

Таким образом, в среднесрочной перспективе речевые технологии и NLP будут развиваться в сочетании с генеративным ИИ. Генеративные модели будут усиливать голосовые системы в сценариях, требующих глубокого понимания контекста и вариативности диалога, тогда как классические речевые и NLP-решения продолжат быть актуальными благодаря меньшей стоимости и ресурсной нагрузке.

Компьютерное зрение (CV)

Развитие технологии CV

Компьютерное зрение³ — это направление искусственного интеллекта, в котором с использованием методов машинного обучения и нейронных сетей компьютеры и технические системы обучаются извлекать осмысленную информацию из визуальных данных и при необходимости автоматически рекомендуют решения или выполняют действия при обнаружении дефектов и аномалий. Исторически технология развивалась более 60 лет, и за это время сформировались несколько значимых вех:

1960–2000-е гг.

Формируются основы: первые методы восстановления трехмерной формы по двумерным изображениям. Наблюдается рост прикладных задач в индустрии и науке.

2001 г.

Появляются системы распознавания лиц в реальном времени.

2010 г.

ImageNet задает стандарт масштабной разметки и оценки качества моделей.

2012 г.

Модель AlexNet на сверточных нейронных сетях резко снижает ошибку распознавания до нескольких процентов — с этого момента глубинные нейросети становятся отраслевым стандартом.

2015–2017 гг.

Развитие идет по траектории снижения задержек инференса, роста пропускной способности при обработке потоков видео и увеличения возможности разворачивать решения не только в облаке, но и на устройствах (на камерах, мобильных и промышленных устройствах) за счет компактных моделей.

2018–2023 гг.

В компьютерном зрении активно применяется архитектура трансформеров, изначально разработанная для обработки естественного языка. В отличие от прежних подходов, где изображение разбивалось на фрагменты и анализировалось локально, трансформеры учитывают связи между удаленными элементами и работают с изображением как с целостной сценой. Это уменьшило зависимость от ручной настройки признаков: модели способны переобучаться на больших массивах неразмеченных изображений и затем дообучаться на небольших размеченных выборках под конкретные задачи, сокращая путь от идеи до прототипа.

Параллельно заметно растет качество генерации изображений и видео, которые становятся заметно более приближенными к реальности. Синтетические материалы становятся источником данных для обучения систем компьютерного зрения: они позволяют наращивать объемы обучающих примеров и целенаправленно воспроизводить редкие или опасные сценарии — аварии, пожары, экстремальные погодные условия — с варьированием обстановки, освещения, ракурса и дистанции. В результате обучение ускоряется и дешевеет, а устойчивость моделей к изменению условий повышается.

2023–2024 гг.

Компьютерное зрение все чаще начинает работать в составе мультимодальных систем, объединяющих изображение и текст. Одна и та же модель одновременно анализирует кадр и интерпретирует текстовый запрос, формируя текстовый ответ. При таком подходе зрение выступает частью единой архитектуры, что позволяет решать прикладные задачи в диалоговом формате — описывать содержимое изображения, находить нужные объекты, отвечать на вопросы по кадру, давать пошаговые инструкции.

Текущий этап характеризуется переходом к парадигме «от восприятия к действию» (vision to action): система по изображению и заданной цели сразу формирует действие, а не только описание. Модель определяет последовательность шагов — взять предмет, открыть дверь, объехать препятствие. Пока такие решения требуют существенной ручной настройки, однако идет движение к полной автоматизации и объединению восприятия и управления в единой модели. Наибольший спрос на подобные подходы наблюдается в робототехнике и механизированном оборудовании.

Ключевые тренды и изменения в технологии СВ

Современное развитие компьютерного зрения опирается на уже описанные архитектуры и ускоряет переход от распознавания к устойчивому действию и интеграции в реальные процессы. В рамках технологии можно выделить следующие ключевые тренды за 2025 г.:

- 3D-пространственное рассуждение. Модель понимает сцену в объеме: по серии фото/видео она восстанавливает форму и размеры объектов, определяет, что где находится, на каком расстоянии, что впереди, а что скрыто за другим предметом. Она строит плотную карту глубины, видит точки соприкосновения и возможные столкновения, по визуальным признакам оценивает материал и физические свойства — где твердая поверхность, где мягкая, где скользко или хрупко. В строительстве и архитектуре это сокращает время съемки и проверок, повышает точность измерений и позволяет быстро сверять фактическое исполнение с BIM/AutoCAD-моделями.
- Прогнозирование движения. Технология позволяет системе зрения не только фиксировать текущее изображение, но и заранее рассчитывать, как оно изменится при следующем повороте или шаге. Раньше восприятие и планирование происходили отдельно: один модуль интерпретировал окружение, другой прогнозировал траекторию маршрута. Теперь одна модель одновременно анализирует текущий кадр и формирует визуальный прогноз следующего кадра. Это превращает камеру в инструмент предиктивного управления и напрямую применяется в автономном транспорте (ADAS/автопилоты, дроны, складские AGV/AMR) для выбора безопасной траектории, удержания дистанции и предотвращения столкновений.
- Реалистичное восприятие света и отражений. Алгоритмы учитывают распространение и многократные отражения света, корректно обрабатывая блики, стекло и полированные поверхности. Это повышает точность восстановления формы и контуров и улучшает сегментацию и измерения в сложных оптических условиях.
- Анализ изображений в исходном качестве. Модель изначально обрабатывает изображения в высоком разрешении до 4К без предварительного сжатия. Раньше изображения необходимо было уменьшить до фиксированного входного размера модели, из-за чего терялись мелкий текст и тонкие контуры. Теперь анализ выполняется в исходном качестве, что повышает точность распознавания деталей и снижает искажения.

- Умное выделение важных деталей. Способ предварительной подготовки данных перед подачей в модель: система автоматически выделяет ключевые фрагменты (текст, номера, мелкие дефекты), анализирует их и объединяет результаты в единый вывод, сохраняя привязку к исходному кадру. Такой подход снижает вычислительные затраты и время отклика и одновременно повышает точность на материалах с большим числом деталей. Технология обеспечивает обработку больших изображений и продолжительных роликов без изменения или переобучения базовой модели.

Основные области внедрения технологии CV

Технология компьютерного зрения прежде всего выступает эффективным инструментом автоматизации визуального контроля, сканирования изображений и отслеживания объектов. Ключевые функции включают: анализ и поиск по изображениям и видео, детекцию и локализацию объектов, выделение ключевых точек и поз для последующего анализа, а также создание новых изображений и видеороликов по текстовому описанию или примерам.

По результатам проведенного опроса, наиболее часто компьютерное зрение применяется на производстве (67%). При этом доля внедрения среди компаний из обычных отраслей выше, чем у компаний из продвинутых в ИИ отраслей (75% против 37%). Это объясняется тем, что в производстве значительная часть операционной деятельности требует контроля с физическим наблюдением. В других же функциях в продвинутых в ИИ отраслях доля использования заметно выше, например в клиентском сервисе (60% против 46%) или в маркетинге и продажах (90% против 29%), где компьютерное зрение используется для распознавания и извлечения структурированных данных, сокращая сроки и ошибки обработки.

В среднем в каждой отрасли компьютерное зрение внедряется в две функции, а в ИТ и технологиях оно применяется больше всего: там среднее количество затронутых функций составляет 3,4. Небольшой разброс в этом показателе среди разных индустрий говорит об универсальной применимости технологии в разных индустриях, но одновременно и об ограниченной масштабируемости по разным функциям.

Доля компаний, внедривших CV в бизнес-функции, от общего числа опрошенных компаний по индустрии, %

	Производство	Клиентский сервис	Маркетинг и продажи	Цепочки поставок	HR	Исследования и разработки	ИТ	Стратегия	Финансы	Внутренние коммуникации	Среднее кол-во функций
IT и технологии	89%	100%	100%			50%					3,4
Строительство и недвижимость	88%	86%	100%								2,7
Банкинг и страхование	20%	100%	80%								2,0
Электроэнергетика	100%	100%									2,0
FMCG	100%		100%								2,0
Ритейл	14%	100%	50%	33%							2,0
Металлы и горная добыча	100%	94%									1,9
Медицина и здравоохранение	80%	100%									1,8
Автомобильная промышленность	100%	75%									1,8
Транспорт и логистика	71%			100%							1,7
E-commerce	40%	80%	50%								1,7
Сельское хозяйство	50%		100%								1,5
Телеком и медиа	40%		100%								1,4
Машиностроение	71%					50%					1,2
Химия и нефтехимия	100%										1,0
Нефть и газ	50%										0,5
Продвинутые в ИИ	37%	60%	90%	13%	13%	0%	0%	0%	0%	0%	2,1
Обычные отрасли	75%	46%	29%	11%	0%	4%	0%	0%	0%	0%	1,7
В среднем	67%	50%	44%	11%	3%	3%	0%	0%	0%	0%	1,8

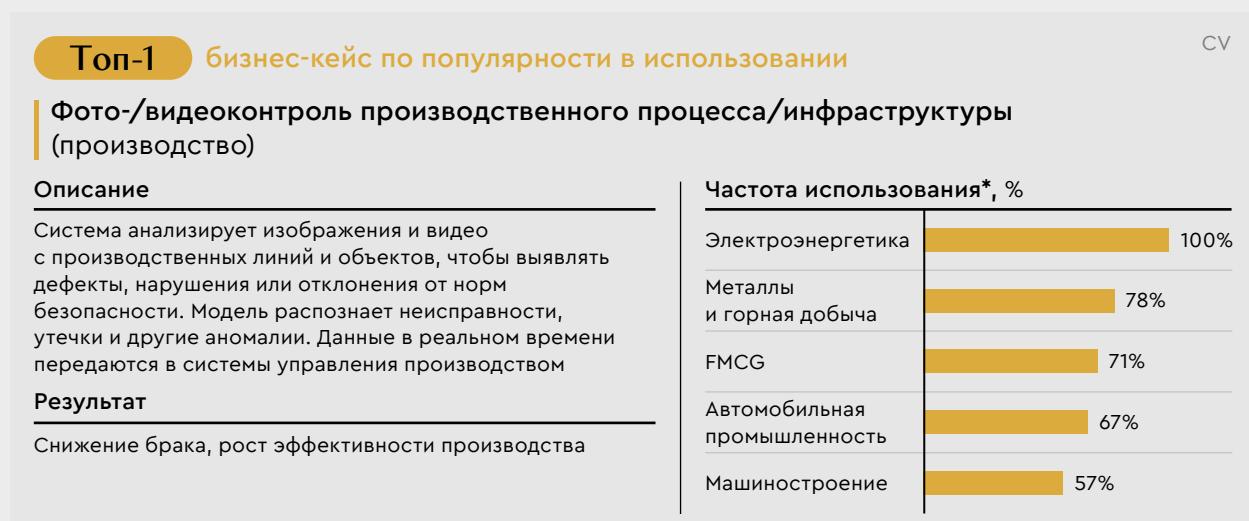
Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.1. В каких функциях внедряется данная технология и в какой степени?)

Примеры практического применения технологий компьютерного зрения охватывают несколько ключевых направлений, каждое из которых находит отражение в конкретных кейсах и решениях.

- Идентификация и анализ человека: распознавание лица (Face ID), а также анализ действий и поведения конкретного пользователя на основе жестов и движений.
- Дополненная реальность (AR): визуализация виртуальных объектов в реальной среде на основе анализа окружения.
- Генерация изображений и видео: создание визуального контента на основе текстовых описаний или эталонных примеров (генеративные модели).
- Распознавание объектов и считывание реквизитов: детекция и извлечение структурированных элементов (например, автоматическое считывание номера банковской карты в мобильном приложении банка, распознавание текстовых документов).

По результатам опроса были выделены топ-5 наиболее распространенных кейсов применения компьютерного зрения в различных отраслях.

Топ-5 бизнес-кейсов использования технологии CV (1/3)



* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

Топ-5 бизнес-кейсов использования технологии CV (2/3)

Топ-2 бизнес-кейс по популярности в использовании

CV

Контроль производительности сотрудников (производство)

Описание

Компьютерное зрение отслеживает действия сотрудников, оценивает скорость и точность выполнения операций, фиксирует нарушения и простои. Модель позволяет выявлять узкие места, оценивать использование рабочего времени и повышать безопасность. Отчеты формируются автоматически и интегрируются с системами производственного планирования

Результат

Повышение производительности и прозрачности

Частота использования*, %

Нефть и газ	50%
Химия и нефтехимия	50%
Строительство и недвижимость	38%
ИТ и технологии	36%
Металлы и горная добыча	33%

Топ-3 бизнес-кейс по популярности в использовании

CV

Поиск товаров по фото (маркетинг и продажи)

Описание

Пользователь загружает изображение товара, а система находит точный аналог или визуально похожие позиции в каталоге. Алгоритм распознает цвет, форму, бренд и подбирает альтернативы из ассортимента. Решение используется как в онлайн-витринах, так и в магазинах самообслуживания

Результат

Ускорение поиска, рост конверсии и среднего чека

Частота использования*, %

E-commerce	43%
Ритейл	41%
FMCG	29%
Телеком и медиа	10%
Прочие индустрии	0%

Топ-4 бизнес-кейс по популярности в использовании

CV, генеративный ИИ

Автоматизация управления складскими запасами с прогнозом потребностей, мониторингом остатков и оптимизацией размещения товаров (цепочки поставок)

Описание

Камеры фиксируют наличие и расположение товаров на складе, подсчитывая единицы хранения. Данные передаются в систему управления запасами, где анализируются с прогнозированием спроса. Система предлагает оптимальное размещение

Результат

Снижение излишков и дефицита, ускорение инвентаризации

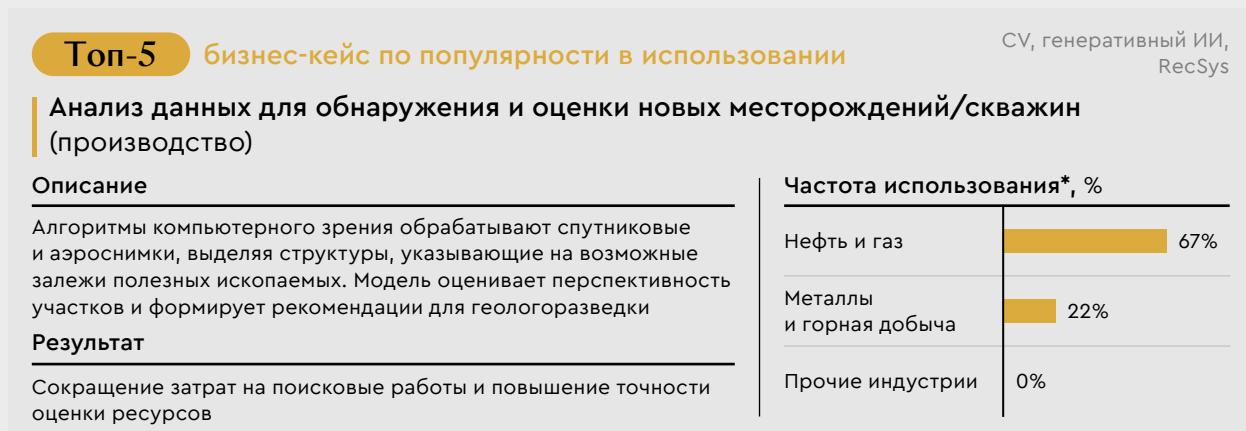
Частота использования*, %

Автомобильная промышленность	33%
E-commerce	29%
Сельское хозяйство	25%
Металлы и горная добыча	11%
ИТ и технологии	8%

* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

Топ-5 бизнес-кейсов использования технологии CV (3/3)



* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

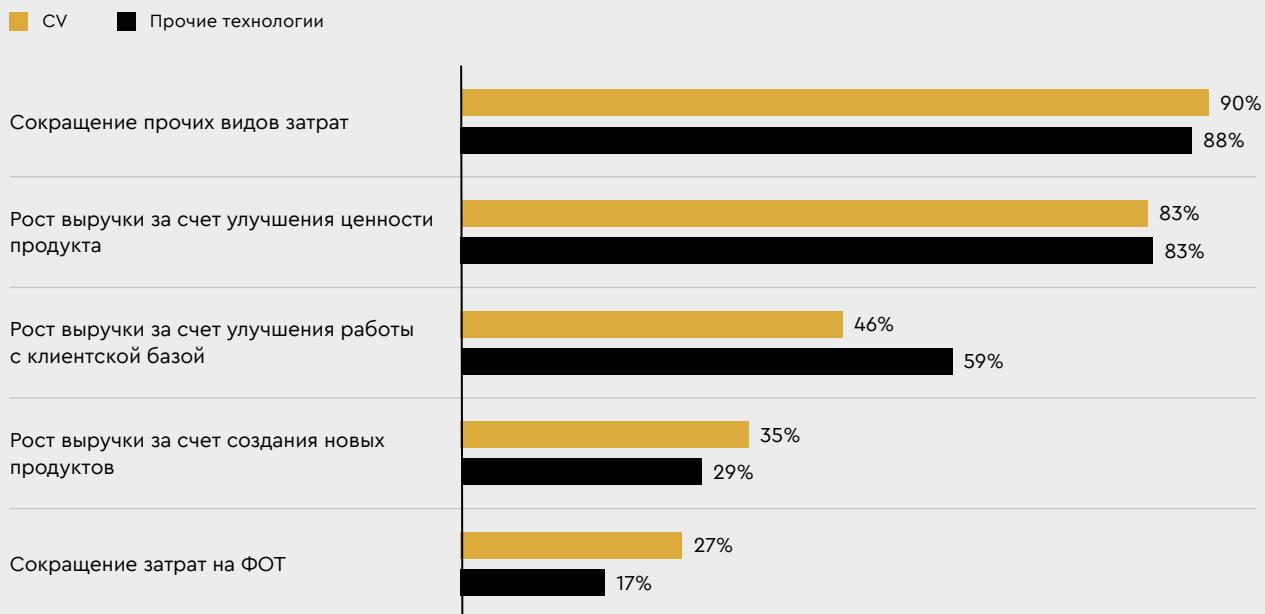
Большинство пользователей (67% респондентов) воспринимают компьютерное зрение в первую очередь как инструмент распознавания лица

При этом, в соответствии с опросом пользователей, ассоциации с CV сегодня сужены до нескольких массовых сценариев — распознавания лиц/оплаты по лицу (69%), видеонаблюдения и безопасности (63%). При этом среди молодых респондентов 18–24 лет заметно чаще упоминаются автономный транспорт и визуальные фильтры, что выделяет их на фоне других возрастных групп. С точки зрения пользователей, CV в основном применяется в ИТ (61%), медицине (50%), телекоме (39%) и банкинге (37%).

Внедрение решений CV в организациях

Большинство респондентов среди компаний отмечают, что ключевой эффект от внедрения технологий компьютерного зрения заключается в сокращении затрат (90%) и росте выручки за счет повышения ценности продукта (83%), что сопоставимо с результатами по другим технологиям (88 и 83% соответственно). Экономия формируется за счет автоматизации визуального контроля: технология снижает трудозатраты на инспекции и одновременно повышает качество, а рост качества, в свою очередь, уменьшает брак и простои. Рост выручки достигается улучшением качества продукта и сервиса (лучшее качество, меньше ошибок). Вместе с тем значительно реже фиксируется рост выручки через улучшение работы с клиентской базой — 46% против 59% по другим решениям. Это отражает отраслевую специфику: в производственных и инфраструктурных сегментах CV прежде всего решает операционные задачи (контроль процесса, безопасности и качества). В то же время в банкинге, ИТ и ритейле технология широко используется в клиентском сервисе (биометрия, KYC). При этом респонденты заметно чаще, чем по другим технологиям, отмечают снижение расходов на ФОТ (27% против 17%) благодаря замещению ручных операций автоматизированным визуальным контролем.

Сложности, которые видят компании в использовании технологий, % ответов (сравнение CV со средним значением по прочим технологиям)



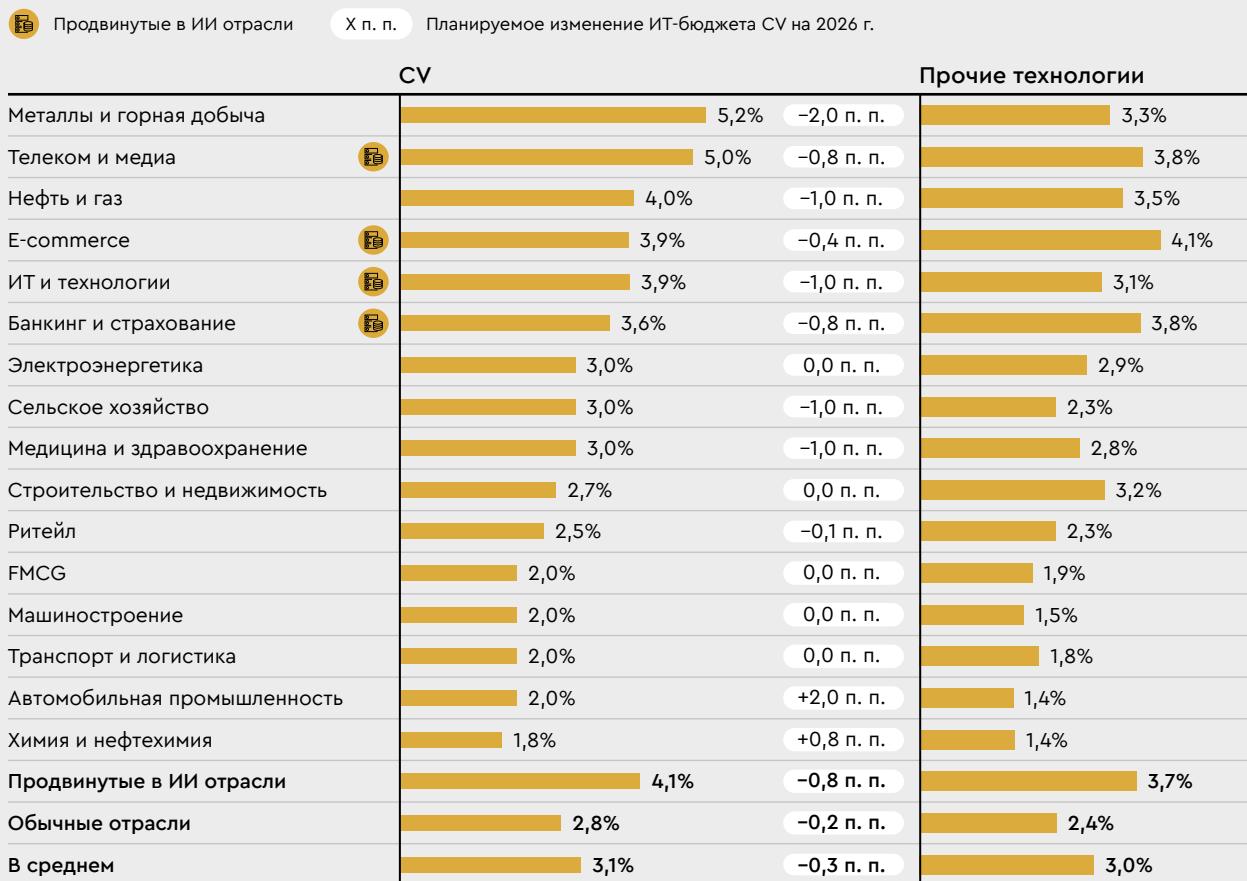
Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.1. В чем компания видит источник финансового эффекта от внедрения технологии?)

Анализ инвестиционной активности показывает, что компании выделяют на компьютерное зрение более высокие бюджеты, чем на другие технологии. В продвинутых в области ИИ отраслях средняя доля инвестиций составляет 4,1% против 3,7% от ИТ-бюджета, в обычных отраслях — 2,8% против 2,4%. Такая разница может быть объяснена совокупностью факторов:

- необходимостью закупки капитальной инфраструктуры (камеры и сенсоры, сеть и хранение видео);
- интеграцией с системами управления производством;
- высокой эффективностью технологии, из-за чего компании склонны тратить большие бюджеты;
- потребностью в настройке каждого индивидуального решения под конкретную ситуацию.

Особенно выделяются отрасли «Металлы и горная добыча» (5,2% от ИТ-бюджета на компьютерное зрение против 3,3% в среднем по другим технологиям) и «Телеком и медиа» (5,0% против 3,8%). В первом случае высокие вложения обусловлены масштабом и рискоемкостью производственных объектов: контроль безопасности, состояния оборудования и качества продукции требует плотной сети камер и аналитики. В телекоме и медиа компьютерное зрение используется для анализа и модерации контента, мониторинга трансляций и персонализации рекламы, что также тянет за собой серьезную инфраструктурную базу. При этом сектор «Металлы и горная добыча» планирует наиболее заметное сокращение инвестиций в 2026 г. — на 2 п. п. от годового ИТ-бюджета (против сокращения на 0,3 п. п. в среднем по рынку и на 0,8 п. п. среди продвинутых в ИИ отраслях). Возможным объяснением служит то, что основные капитальные вложения уже сделаны и фаза масштабирования сменяется этапом оптимизации существующих решений и повышения отдачи от инвестиций — с меньшими дополнительными вложениями.

Средние инвестиции компаний в технологии ИИ за последний год, в % от годового ИТ-бюджета (сравнение CV со средним значением по прочим технологиям)



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.3. Какой объем средств в % от годового ИТ-бюджета компания уже инвестировала в технологию за последний год? Вопрос 4.5. Какой объем средств в % от ИТ-бюджета компания планирует инвестировать во внедрение технологии на горизонте года?)

В большинстве случаев технологии компьютерного зрения разрабатываются с нуля, что существенно усложняет внедрение и увеличивает стоимость технологий

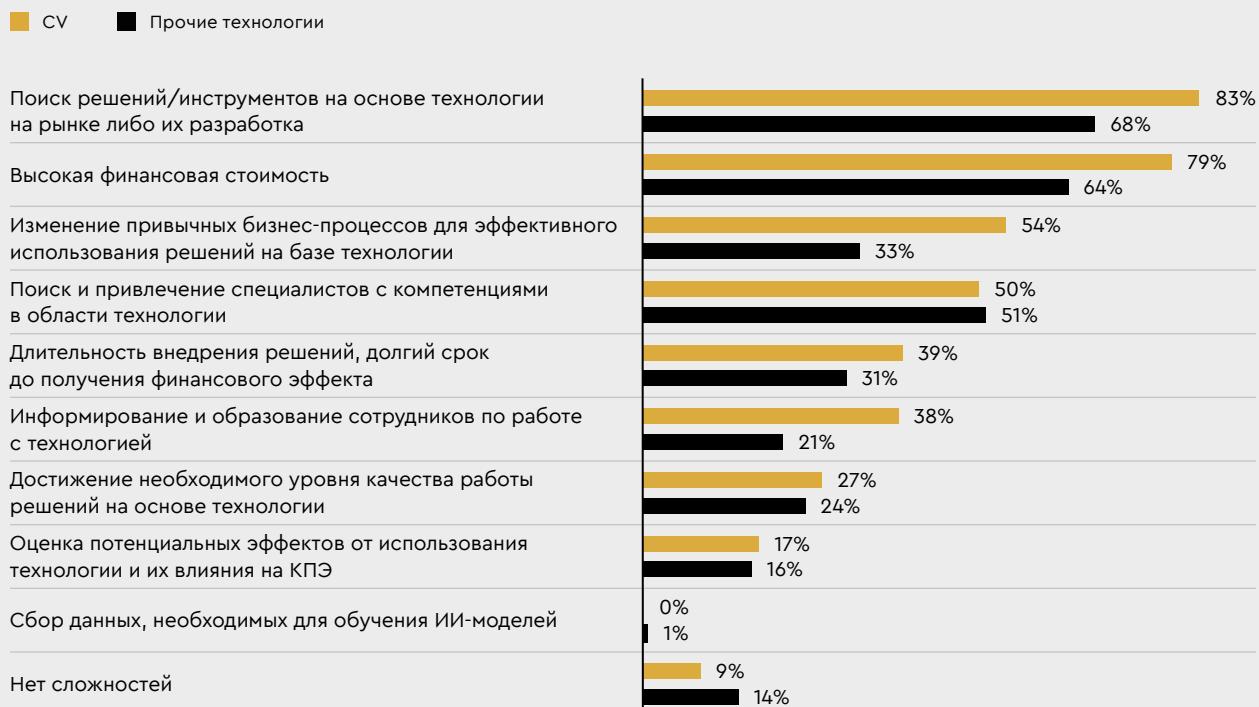
Высокая стоимость и капиталоемкость не только увеличивают бюджет, но и выступают основным барьером внедрения: на это указывают 79% респондентов (против 64% в среднем по другим технологиям). Помимо цены, чаще других отмечаются:

- поиск зрелых решений под специфичные кейсы и сложность выбора вендора (83% против 68% в среднем по другим технологиям);
- необходимость перестройки бизнес-процессов и ответственности за результат (54% против 31%);
- переобучение персонала для работы с новой технологией (38% против 21%).

Обычные пользователи не сильно осведомлены о технологиях CV (знают — 27%, глубокое понимание — у 6%).

Все это подтверждает факт того, что внедрение конкретных кейсов компьютерного зрения в функции бизнеса является сложным процессом, а масштабирование затруднено из-за специфики выполняемых задач. Более того, как утверждают вендоры решений компьютерного зрения, технология всегда реализуется на базе разработанных моделей с нуля из-за особенностей функционала каждого решения компьютерного зрения.

Сложности, которые видят компании в использовании технологий, % ответов (сравнение CV со средним значением по прочим технологиям)



Источник: опрос индустриальных компаний (блок 5. Барьеры и сложности при внедрении технологий, вопрос 5.2. Какие сложности видят компания в использовании технологий?)

С другой же стороны, общее отношение к компьютерному зрению в обществе преимущественно позитивное: 84% респондентов настроены положительно, 57% видят в технологии больше пользы, чем рисков (наибольший результат среди ключевых ИИ-технологий). Особенно высока поддержка среди специалистов по ИТ (82%). Такое социальное восприятие технологии может быть существенным преимуществом для компаний в масштабировании сервисных сценариев при соблюдении норм приватности и объяснимости решений.

Более того, несмотря на капиталоемкость и сложность реализации, компании оценивают влияние компьютерного зрения на EBITDA как наиболее высокое среди сравниваемых технологий: средняя величина эффекта — 1,0% EBITDA против 0,8% по другим решениям. Такое происходит из-за высокой эффективности технологии в тех функциях, где она применяется. При этом наибольший эффект, как и в других технологиях, наблюдается в продвинутых в ИИ отраслях: все они входят в топ-6 по влиянию на EBITDA. При этом отрасль «Металлы и горная добыча», несмотря на наибольшую долю инвестиций в СУ, занимает лишь четвертое место по достигнутому эффекту. Это объясняется спецификой применения технологии в секторе:

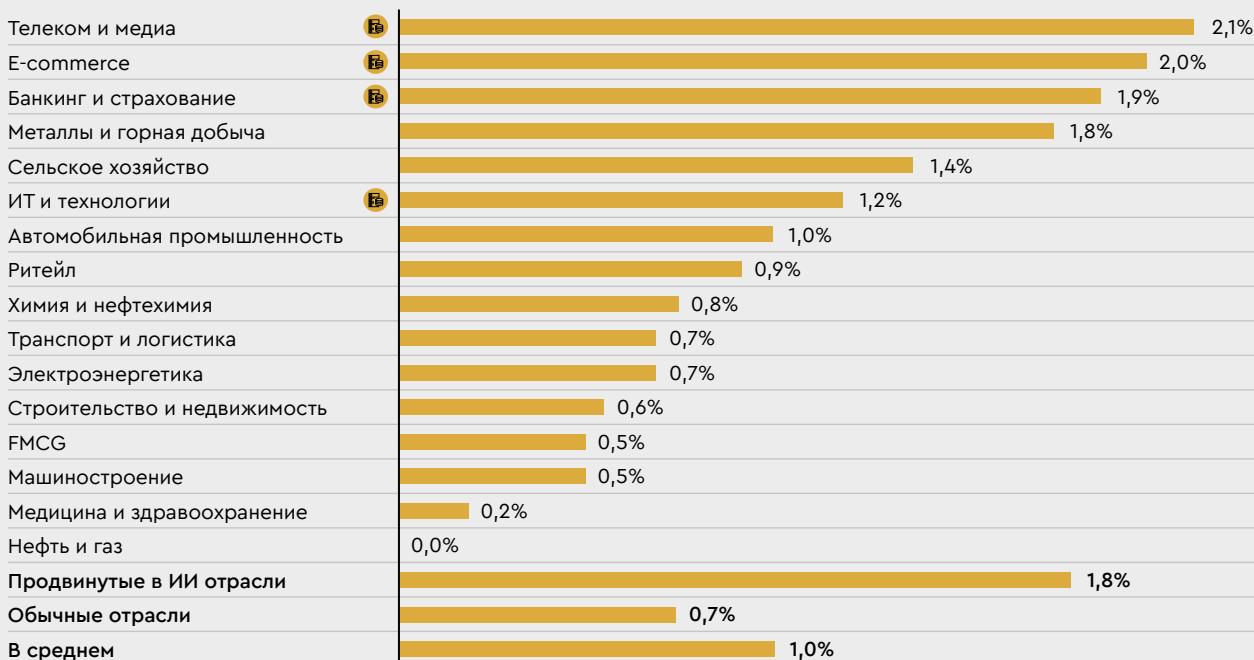
- доминируют кейсы мониторинга активов (опасные зоны, состояние конвейеров, обнаружение дыма/пламени), где эффект выражается прежде всего в снижении рисков, инцидентов и штрафов, а не напрямую в росте выручки;
- эффект имеет «потолок»: после покрытия ключевых площадок камерами и аналитикой дальнейшее улучшение дает убывающую отдачу (идет оптимизация вместо масштабного наращивания);
- разрозненная география объектов и ограничения связи ведут к сложной интеграции с ОТ/SCADA и более длинным циклам внедрения и измерения эффекта.

В результате сектор переходит от фазы интенсивных внедрений к этапу оптимизации существующих решений и повышению отдачи от уже сделанных инвестиций, что и отражается в текущей динамике по эффекту.

Средний эффект компаний от внедрения компьютерного зрения за последний год, % EBITDA



Продвинутые в ИИ отрасли



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.2. Какой финансовый эффект компания уже получила с момента внедрения технологии в % от годовой EBITDA за последний год?)

Рынок B2B-решений технологий CV

Проведенный опрос показал, что среди российских вендоров, принявших участие в исследовании, 21% специализируются на технологиях компьютерного зрения. Это значение является вторым после вендоров ИИ-помощников, доля которых составляет 34%.

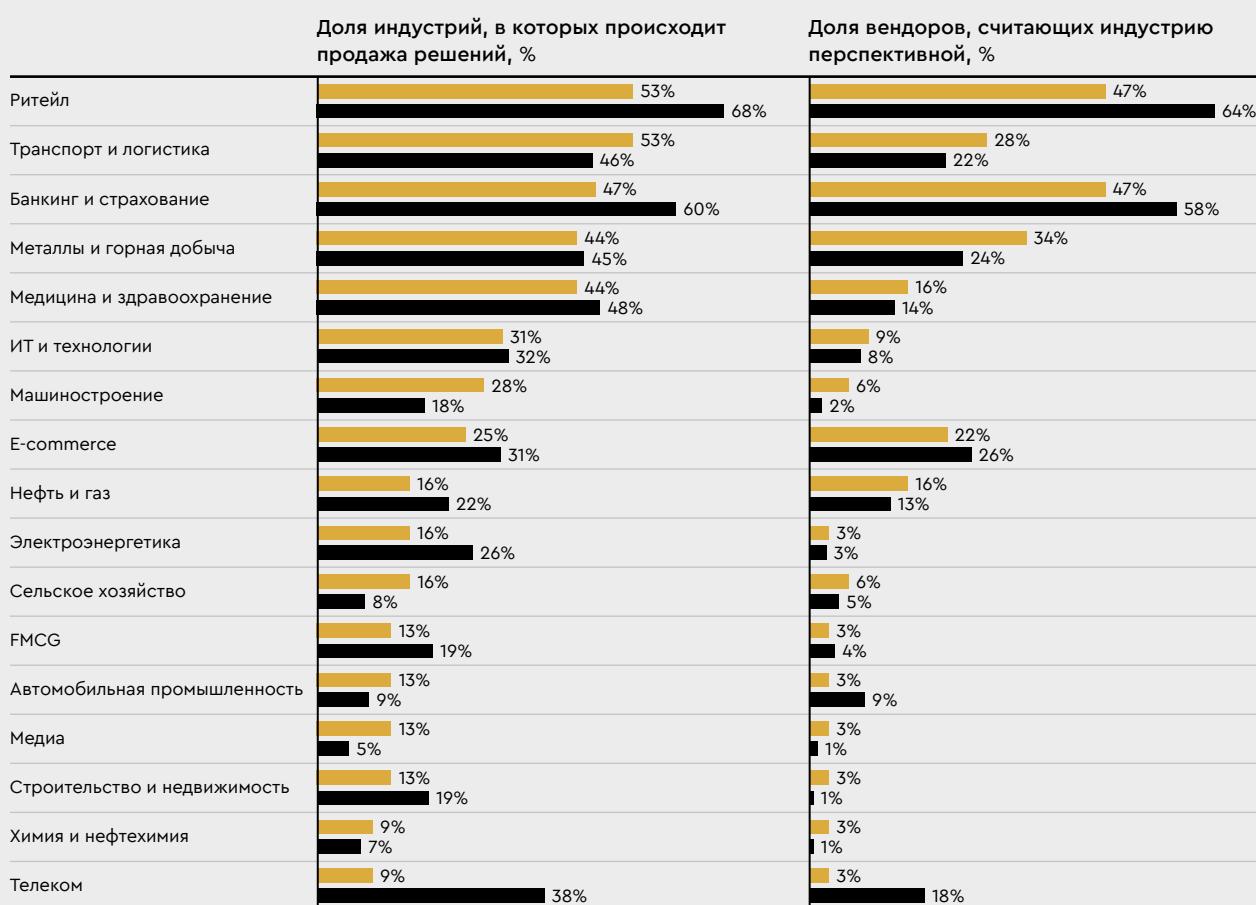
94% опрошенных вендоров, специализирующихся на компьютерном зрении, присутствуют на рынке более 5 лет, что является наивысшим показателем среди всех технологий ИИ

Рынок B2B-решений на основе компьютерного зрения характеризуется рядом особенностей. 94% опрошенных вендоров присутствуют на рынке более пяти лет — это наибольший показатель зрелости среди всех вендоров технологических направлений. Это связано с тем, что технология развивается уже на протяжении значительного времени и требует не просто внедрения готовых решений, а создания моделей с нуля, их глубокой интеграции в производственные процессы и обеспечения высокой надежности. При этом рынок представлен компактными командами: у 87% вендоров в штате менее 200 сотрудников — это наибольшая доля небольших компаний среди сравниваемых технологий. Несмотря на компактные размеры, 16% компаний, специализирующихся на компьютерном зрении, имеют годовую выручку свыше 1 млрд руб., что существенно выше среднего уровня по остальным технологиям (6%). Такая динамика отражает высокую сложность технологии: решения в области компьютерного зрения требуют кастомизации под отрасль, конкретный объект/кейс и имеют сравнительно низкую частоту продаж, что обуславливает высокую стоимость проектов и услуг вендоров.

Профиль клиентов компаний, работающих в сфере компьютерного зрения, в целом сопоставим с профилем заказчиков других технологических направлений. Ключевые отрасли применения — ритейл (53% против 68% среди прочих технологий), в котором типовые кейсы применения связаны с маркетингом и управлением запасами (поиск товара по фото, автономная инвентаризация стеллажей на складе), а также транспорт и логистика (53% против 46% среди прочих технологий). В структуре отраслей, которые вендоры считают перспективными, выделяется сектор «Металлы и горная добыча»: 34% компаний, специализирующихся на компьютерном зрении, считают его перспективным против 24% среди вендоров других технологий. Это объясняется большим количеством прикладных задач и масштабами инфраструктуры: мониторинг конвейеров, подсчет и трекинг техники, визуальная диагностика оборудования, контроль отвалов и хвостохранилищ по аэросъемке/спутникам. Для таких кейсов требуется плотная сеть камер и аналитика, что формирует устойчивый спрос на решения компьютерного зрения.

Индустрии, в которых происходят продажи у вендоров/доля вендоров, считающих индустрию перспективной, % ответов (сравнение CV со средним значением по прочим технологиям)

■ CV ■ Прочие технологии



Источник: опрос вендоров (блок 4. Портфель клиентов, вопрос 4.1. В каких индустриях компания продает решения на основе технологии и какие считает наиболее перспективными?)

Будущее и перспективы развития технологии компьютерного зрения

На текущем этапе развития технологии компьютерного зрения в России большинство компаний оценивают ее уровень как выше среднего по сравнению с мировыми показателями — как в части разработки, так и в части внедрения.

- По уровню разработки: 93% компаний-респондентов отметили, что российские решения находятся выше среднего мирового уровня или на уровне стран-лидеров в области ИИ. По этому показателю компьютерное зрение занимает второе место среди всех технологий, уступая лишь рекомендательным системам (RecSys), у которых аналогичный показатель составляет 99%.
- По уровню внедрения: 87% компаний-респондентов считают, что технологии компьютерного зрения применяются в России на среднемировом или более высоком уровне, что также является вторым результатом среди рассматриваемых направлений (после RecSys — 98%).

Более того, пользовательские ожидания коррелируют с оценками компаний: 77% респондентов верят в расширение применения СВ, а каждый пятый (20%) считает, что технология станет ключевой во многих бизнес-процессах. При этом 57% оценивают позицию России выше среднего, а среди ИТ-аудитории доля положительных оценок достигает 64%.

При этом 37% опрошенных вендоров и 35% компаний-респондентов ожидают, что в ближайшие годы произойдет качественный скачок в развитии технологий компьютерного зрения, что приведет к появлению новых решений и сценариев их применения.

- Рetail и логистика: магазины без касс, контроль выкладки в реальном времени, мониторинг складов.
- Строительство и инфраструктура: массовая 3D-реконструкция и цифровые двойники объектов, автоматическая съемка, контроль отклонений, объемов и качества работ.
- Медицина: автоматизированный анализ биопсии — выделение опухолевых областей, подсчет клеток/митозов, оценка биомаркеров — с формированием подсказок врачу и контролем качества протокола.

Кроме того, 83% вендоров и 63% компаний прогнозируют резкий или постепенный рост использования технологий компьютерного зрения.

Дополнительным драйвером становятся VLA-модели (vision-language-action models), которые объединяют восприятие компьютерного зрения и языковое рассуждение генеративных моделей в едином контуре: наблюдение, понимание, действие.

Такие модели умеют связывать объекты сцены с их свойствами и возможностями, строить планы и выполнять многошаговые инструкции в реальном мире и симуляторах. На практике это открывает путь к робототехническим ассистентам, автономной инспекции и навигации на промышленных площадках, а также к AR-интерфейсам, где подсказки и действия формируются по контексту камеры в режиме реального времени.

Интеграция компьютерного зрения с генеративным ИИ выходит за рамки генерации контента: генеративный ИИ выступает источником синтетических данных и редких сценариев для обучения, модулем объяснимого рассуждения при разборе сложных сцен и мозгом для составления пошаговых планов действий.

Таким образом, синергия компьютерного зрения и генеративного ИИ формирует новый класс интеллектуальных систем, способных не только фиксировать и интерпретировать визуальные данные, но и обобщать и прогнозировать нетипичные условия, а также адаптироваться к ним.

Вместе с тем остаются факторы, ограничивающие масштабирование и широкое применение компьютерного зрения:

- Высокая стоимость подготовки данных. Обучение систем CV требует анализа значительных объемов визуальной информации и частично ручной разметки изображений и видео. Несмотря на использование синтетических данных, потребность в тщательно структурированных наборах остается высокой, что увеличивает затраты и продолжительность обучения по сравнению с языковыми моделями, работающими с неразмеченными текстами.
- Отсутствие универсальных метрик. Эффективность CV-моделей невозможно оценивать по единому критерию: показатели точности и устойчивости зависят от отрасли и конкретных бизнес-задач, что затрудняет сопоставимость решений.
- Ограниченностю обычных камер. Без использования дополнительных датчиков (LiDAR, IMU) стандартные камеры не обеспечивают достаточной точности 3D-восстановления, что снижает надежность при автономных действиях.
- Снижение устойчивости в непредусмотренных ситуациях. При столкновении с объектами, отсутствующими в обучающих данных, модели могут ошибочно их классифицировать, что создает риски при эксплуатации в реальных условиях.

В совокупности эти факторы подтверждают, что, несмотря на высокий потенциал и активное развитие, компьютерное зрение остается технологией, требующей дальнейшего совершенствования инфраструктуры, алгоритмов и стандартов качества.

Рекомендательные системы (RecSys)

Развитие технологии RecSys

1990-е гг.

Коллаборативная фильтрация становится отправной точкой развития рекомендательных систем: в начале 1990-х гг. в Xerox PARC система Tapestry⁴ фильтрует корпоративную почту по меткам и отзывам коллег, а вскоре подход переносят на потоки сообщений (например, GroupLens для Usenet⁵). Базовый принцип прост: если пользователи с похожими интересами положительно оценили объект, его стоит рекомендовать другим похожим пользователям. Ранние реализации опираются на явные оценки и соседские методы сходства, решая задачу отбора релевантных писем и сообщений и тем самым закладывая методическую основу всей последующей эволюции рекомендаций.

2000-е гг.

Рекомендательные механики начинают встраиваться в поисковые модели: подсказки запросов и похожие результаты становятся стандартом крупных поисковых систем. Ключевой вехой становится запуск Google Personalized Search⁶ в 2005 г.: результаты начинают перестраиваться с учетом истории действий пользователя. С этого момента поиск и рекомендации эволюционируют как единая система, где поведенческие сигналы и модели персонализации влияют и на порядок выдачи, и на навигацию по результатам.

2010-е гг.

Рекомендательные системы переходят к каскадной схеме «отбор кандидатов — ранжирование» (YouTubeDNN⁷, Wide&Deep, DeepFM): сначала из огромного каталога быстро выбираются десятки подходящих позиций, а затем они упорядочиваются под цели витрины. Ключевой сдвиг — умение работать со множеством признаков о пользователе, объекте и контексте, включая признаки с миллионами и миллиардами уникальных значений (например, идентификаторы). Вместо ручной сборки сочетаний признаков модели сами учатся находить важные взаимодействия: редкие идентификаторы кодируются векторными представлениями, более высокие зависимости извлекаются внутри сети. На шаге отбора используются парные представления пользователя и объекта, на шаге ранжирования поведенческие, контекстные и товарные признаки объединяются с выученными.

Итог — заметный прирост точности без взрывного роста ручной настройки и масштабируемая персонализация на очень больших каталогах. При этом важно заметить, что при отслеживании действий данные обрабатываются агрегированно и хранятся ограниченно: система анализирует только поведение в самом продукте.

2018–2019 гг.

На первый план выходит не только релевантность каждого элемента по отдельности, но и качество ответа целиком. Модели учатся собирать сбалансированные подборки — плейлисты, ленты, выдачи, — где одновременно учитываются разнообразие, покрытие разных интересов пользователя, новизна/«приятная неожиданность» и уместный порядок.

2021–2022 гг.

Начинает развиваться графовый подход, учитывающий связи между пользователями, товарами и контекстами (совместные покупки, последовательные просмотры). На практике его часто комбинируют с последовательными и контентными моделями, получая дополнительный прирост качества. При этом сегодня графовые методы остаются преимущественно исследовательским направлением: их активно изучает академическое сообщество, но в большинстве отраслей они применяются редко и пока не стали индустриальным стандартом. Среди публично известных крупных внедрений — Pinterest (графовые рекомендации пинов) и Twitter/X⁸, где графы встроены в общий стек, а не заменяют его.

2024–2025 гг.

На уровень надстройки выходят большие языковые модели: они интерпретируют сложные запросы на естественном языке, уточняют намерения, формируют объяснения и могут управлять выдачей. Пока интеграция LLM преимущественно развивается в академических и исследовательских проектах. Компании же проводят точечные пилоты. В дальнейшем планируется прикладное применение в конкретных бизнес-кейсах.

Только 11% пользователей хорошо понимают, как функционируют рекомендательные системы

Эволюция выглядит непрерывной: от коллаборативной фильтрации и ранних интеграций с поиском — к каскадным глубинным моделям, затем к графовым методам, а сегодня — к генеративным и языковым технологиям, которые делают рекомендации объяснимыми, интерактивными и управляемыми при сохранении инженерной эффективности. В то же время пользовательская перспектива задает важную планку: хотя 48% респондентов знакомы с самим понятием рекомендательных систем, лишь 11% хорошо понимают, как они работают, поэтому требования к объяснимости и диалоговым сценариям становятся частью базового UX.

Ключевые тренды и изменения в технологии RecSys

За последние годы рекомендательные системы претерпели заметную трансформацию. Их развитие тесно связано с переходом от узкоспециализированных алгоритмов к более универсальным архитектурам. Компании стремятся сделать рекомендации не только точными, но и прозрачными, объяснимыми и доступными в реальном времени. На текущем этапе выделяются несколько ключевых векторов развития:

- Масштабирование как драйвер качества. Подтверждены законы масштабирования: рост объема данных, параметров и вычислений дает предсказуемое улучшение метрик (конверсия, удержание клиента). Это закрепляет курс на крупные базовые модели и позволяет планировать траектории качества и бюджет.
- Отказ от каскадов в пользу единой модели. Переход от жесткой схемы «кандидаты — ранжирование» к единым генеративным трансформерным моделям, которые совместно определяют отбор, переупорядочивание и формирование списка, учитывая цели продукта и ограничения монетизации в одном проходе.
- Новый виток применения RL к рекомендательным системам. Обучение с подкреплением (RL) возвращается как способ оптимизировать не отдельный клик, а долгосрочную ценность: удержание, LTV, качество сессии. В рекомендациях вознаграждение верифицируемо (показы, клики, конверсии), что способствует масштабированию RL.
- Мультидоменные, мультимодальные, экосистемные рекомендации. Рекомендательная логика строится как единое ядро персонализации для разных доменов и продуктов одной экосистемы. Система объединяет признаки текста, изображений, видео, логов и метаданных в общее представление «пользователь — объект», чтобы использовать их на этапах отбора кандидатов и последующего ранжирования. За счет кросс-продуктового профиля предпочтения переносятся между сервисами (поиск, лента, магазин, подписки, финансовые сервисы), что повышает точность, разнообразие и устойчивость к холодному старту и дает согласованную выдачу во всех каналах.

- Объяснимость рекомендаций. Системы будут объяснять, почему именно этот товар или контент предложен. Это повышает доверие и упрощает контроль качества в чувствительных сферах, таких как финансы, медицина, образование.
- Диалоговые сценарии. Система общается как консультант: уточняет запрос, предлагает альтернативы, перестраивает список в процессе диалога. Такой интерфейс ускоряет путь к релевантному выбору и упрощает процесс подбора для клиента.

В совокупности эти векторы формируют новую операционную норму для рекомендательных систем: единое ядро персонализации на масштабе данных и вычислений, управляемое целевыми метриками и поддержанное объяснимостью и диалоговыми интерфейсами. В результате рекомендации переходят от статичных списков к интерактивным сценариям, обеспечивая более предсказуемый прирост качества при сохранении баланса между точностью, скоростью и требованиями контроля.

Основные области внедрения технологии RecSys

На основании проведенного опроса можно увидеть, что функциональный профиль внедрения рекомендательных систем имеет устойчивый приоритет клиентских направлений. Наиболее частое использование фиксируется в маркетинге и продажах, где эту технологию применяют около 77% компаний, а также в клиентском сервисе, где эта доля составляет около 69%. В продвинутых по ИИ отраслях технология применяется, как правило, в трех бизнес-функциях, в обычных — преимущественно в двух. Лидирующие позиции по масштабу охвата демонстрируют ИТ и технологии, химия и нефтехимия, а также электроэнергетика, где решения задействованы в среднем в 4,4, 3,5 и 3 функциях соответственно.

Например:

- В ИТ и технологиях рекомендательные системы применяются как во внешних, так и во внутренних бизнес-функциях. На клиентской стороне RecSys используется для персонализации цифровых продуктов и сервисов: персонализация ленты, подбор фильмов/сериалов и музыки в стримингах, а также рекламные рекомендации. Внутри компаний технологии применяются для поддержки сотрудников: подбор курсов и карьерных треков, автоматические рекомендации по решению ИТ-запросов и устраниению инцидентов в сервис-десках.
- В химии и нефтехимии RecSys используют для комплектации заказов и рекомендации продуктовых линеек под ТЗ клиента (добавки, упаковка, партии), динамического ценообразования и скидок по контрактам, а также для персонализации технической поддержки и обучающих материалов.
- В электроэнергетике — для подбора тарифов и сервисных пакетов по профилю потребления, приоритизации предложений по энергосбережению и умному учету, а также для маршрутизации обращений и планирования коммуникаций при ремонтах и отключениях.

На базе рекомендательных систем формируются решения, которые основаны на большом наборе вариантов выбора, то есть на тех вариантах, которые с наибольшей вероятностью окажутся полезными конкретному пользователю в текущий момент. Типовой принцип работы строится на отборе подходящих действий и их ранжировании с учетом определенных правил. Примеры применения можно найти в любых задачах выбора — от витрин и чатов до рассылок и ценообразования. По результатам опроса были выделены топ-5 наиболее распространенных кейсов применения рекомендательных систем в различных отраслях.

Доля компаний, внедривших RecSys в бизнес-функции, от общего числа опрошенных компаний по индустрии, %

	Маркетинг и продажи	Клиентский сервис	Производство	HR	ИТ	Цепочки поставок	Стратегия	Финансы	Исследования и разработки	Внутренние коммуникации	Среднее кол-во функций
ИТ и технологии		100%	83%	86%	75%	100%					4,4
Химия и нефтехимия		100%	67%	80%	100%						3,5
Электроэнергетика			100%	100%	100%						3,0
Строительство и недвижимость		92%	100%			100%					2,9
Транспорт и логистика		100%		86%				100%			2,9
Автомобильная промышленность		100%	83%	100%							2,8
Телеком и медиа		100%	100%	25%	20%	33%					2,8
Медицина и здравоохранение		100%	100%	75%							2,8
E-commerce		100%	75%		33%			33%			2,4
Сельское хозяйство		100%	100%	25%							2,3
Металлы и горная добыча		33%	88%	100%							2,2
Ритейл		100%	100%								2,0
Банкинг и страхование		100%	60%			25%					1,9
FMCG		100%		50%							1,5
Машиностроение			50%	57%							1,1
Нефть и газ				25%							0,3
Продвинутые в ИИ		100%	80%	28%	32%	40%	8%	0%	0%	0%	2,9
Обычные отрасли		69%	66%	58%	25%	0%	8%	0%	0%	0%	2,3
В среднем		77%	69%	51%	27%	10%	8%	0%	0%	0%	2,4

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.1. В каких функциях внедряется данная технология и в какой степени?)

Топ-5 бизнес-кейсов использования технологии RecSys (1/2)

Топ-1 бизнес-кейс по популярности в использовании

Генеративный ИИ, речевые технологии и NLP, RecSys

Автоматизированный чат-бот клиентской поддержки (клиентский сервис)

Описание

Чат-бот на основе ИИ принимает обращения клиентов, отвечает на типовые запросы, помогает решать распространенные проблемы. Система интегрируется с базами знаний, CRM и внутренними системами компании, обеспечивая быструю и точную обработку запросов без участия оператора

Результат

Сокращение нагрузки на поддержку и повышение скорости обслуживания

Частота использования*, %



Топ-2 бизнес-кейс по популярности в использовании

RecSys

Сегментация клиентов (маркетинг и продажи)

Описание

Клиентская база разбивается на группы по ценности и намерениям. Помимо классических признаков, используются также поведенческие представления, отражающие сходство траекторий действий. Сегменты регулярно обновляются и автоматически передаются в CRM и платформы маркетинга

Результат

Повышение отклика на кампании и снижение доли нерелевантных коммуникаций

Частота использования*, %



Топ-3 бизнес-кейс по популярности в использовании

RecSys, генеративный ИИ

Персонализированные предложения клиентам (маркетинг и продажи)

Описание

На главной странице, в категориях, карточке товара и в личном кабинете система предлагает позиции с максимальной вероятностью интереса. В расчет включаются просмотры и их длительность, добавления в корзину, покупки, возвраты, а также их наличие, маржинальность и ограничения по показам

Результат

Рост конверсии, среднего чека и перекрестных продаж

Частота использования*, %



* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

Топ-5 бизнес-кейсов использования технологии RecSys (2/2)

Топ-4 бизнес-кейс по популярности в использовании

RecSys, генеративный ИИ

Персонализация маркетингового контента по поведению, профилю и контексту (маркетинг и продажи)

Описание

Подбираются креатив, тема письма, баннер, канал и время отправки. Используются сигналы профиля, свежие действия, источник трафика, устройство и время суток

Результат

Повышение открытия, кликов и целевых действий без увеличения частоты коммуникаций

Частота использования*, %



Топ-5 бизнес-кейс по популярности в использовании

RecSys

Рекомендации по ценам с учетом рынка, конкурентов и истории поведения (маркетинг и продажи)

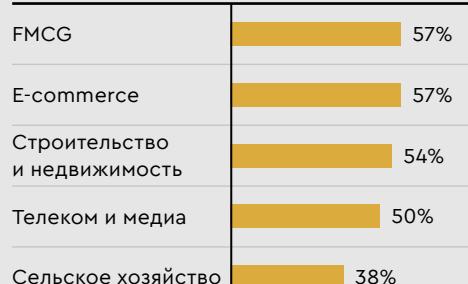
Описание

Для товара или категории выбирается оптимальная цена внутри допустимого коридора. Входные данные включают конкурентов, остатки, сезонность, эластичность спроса и реакцию клиентов на предыдущие изменения. Модель ранжирует ценообразование по различным критериям

Результат

Более стабильная маржинальность и оборачиваемость при меньшем объеме ручных операций

Частота использования*, %



* Частота использования кейса в индустрии отражает его распространенность в целом, включая случаи реализации на базе других технологий

Источник: опрос индустриальных компаний (блок 2. Область внедрения технологии, вопрос 2.2. Какие решения на основе технологии внедряются в данной функции?)

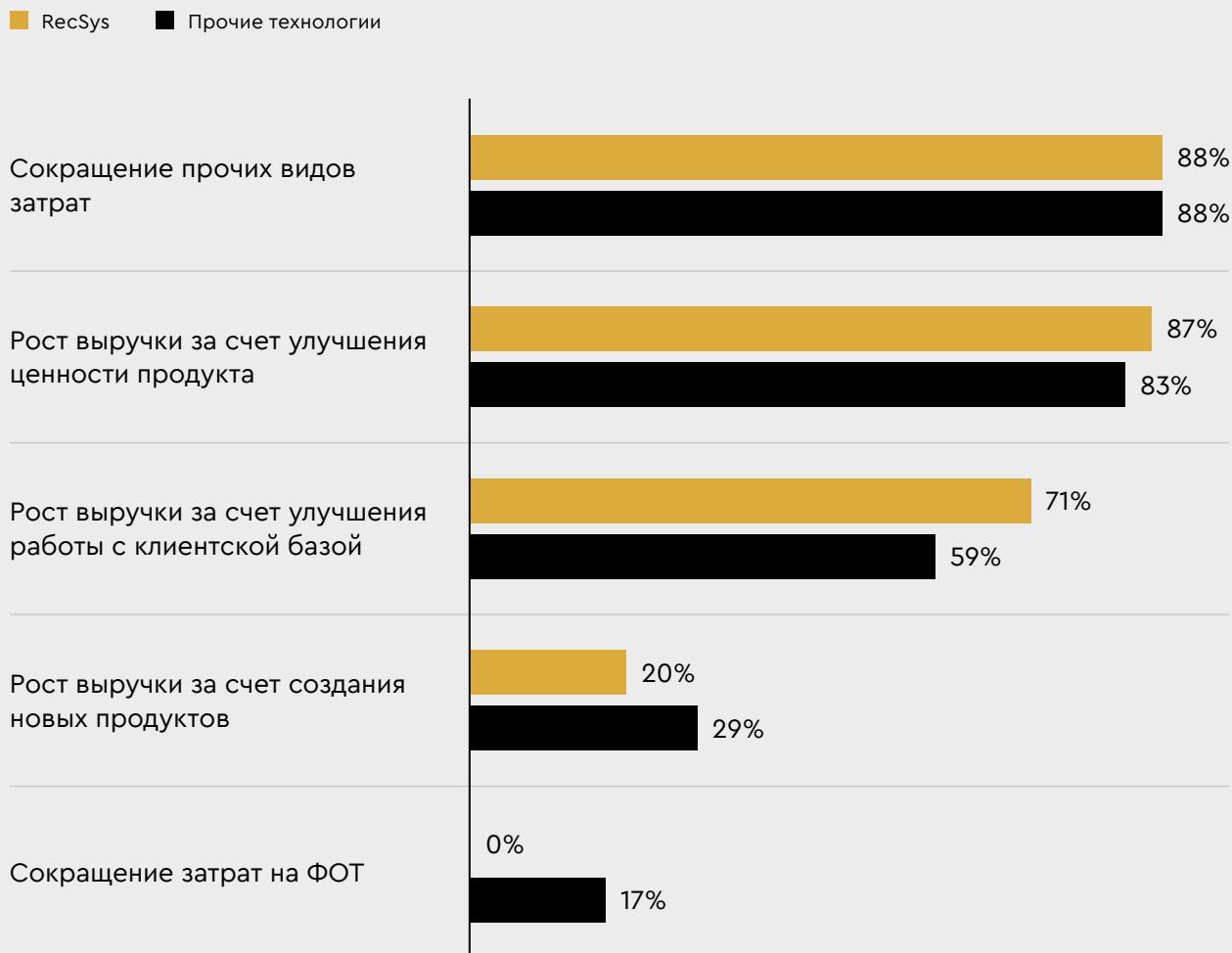
Единая технологическая база позволяет переносить решения между каналами и поверхностями. Измеряемый прирост складывается из небольших улучшений в каждой точке контакта и отражается в росте выручки, удержания и качества сервиса.

С точки зрения пользователей, главное достоинство технологии — облегчение поиска и экономия времени: это отметили 54% респондентов, что стимулирует развитие быстрых подборок и диалоговых помощников.

Внедрение решений RecSys в организациях

Для большинства организаций одним из ключевых итогов внедрения рекомендательных систем в сравнении с другими технологиями выступает повышение эффективности работы с клиентской базой: этот эффект отмечают более 70% респондентов, тогда как по другим технологическим направлениям его указывают менее 60%. При этом из-за специфики кейсов использования рекомендательных систем ни один респондент не выделил сокращение затрат на ФОТ как эффект от технологии. Другие эффекты выделяются сопоставимо с прочими технологиями.

**Эффект от внедрения технологии, ответы компаний в РФ,
% (сравнение RecSys со средним значением по прочим технологиям)**



Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.1. В чем компания видит источник финансового эффекта от внедрения технологии?)

E-commerce инвестирует 4% ИТ-бюджета в рекомендательные системы, что является наибольшим показателем в сравнении с другими отраслями

Инвестиционная картина в RecSys подтверждает зрелость технологии. Доля корпоративных ИТ-бюджетов, направляемых на рекомендательные системы, в среднем ниже (1,7% против 2,7%), что объясняется низкой стоимостью внедрения и тем, что значимая часть капитальных вложений уже произведена в предыдущие годы. На этом фоне в 9 из 16 отраслей заявлены планы сокращения расходов, чаще всего в формате удержания действующих систем и точечных улучшений без масштабных программ модернизации. При этом речь идет не о снижении приоритета технологии, а скорее о переходе к режиму эксплуатационной эффективности и адресной донастройки под конкретные функции.

Среди долей рекомендательных систем в ИТ-бюджетах компаний выделяется e-commerce, где доля бюджета выше среднего и достигает около 4%. Для этого сегмента персонализированные подборки на витрине, в поиске, карточках товаров и в коммуникациях являются критически важными элементами коммерческой модели, поэтому инвестиции поддерживаются на повышенном уровне. По данным опроса, именно в e-commerce наблюдается один из наибольших экономических эффектов от рекомендательных систем за 2024 г., близкий к 2% EBITDA, тогда как в других отраслях среднее значение оценивается на уровне около 0,9%.

В секторах ИТ и технологий, телекоммуникаций и медиа сохраняется сравнительно низкая доля ИТ-бюджетов, направляемых на рекомендательные системы. Это может объясняться крупным абсолютным объемом ИТ-расходов: при невысокой стоимости технологии ее относительная доля оказывается ниже, чем в других отраслях. Несмотря на это, компании отмечают, что эффект от внедрения выше среднего — на уровне 1,5–1,9% EBITDA.

Средние инвестиции компаний в технологии ИИ за последний год, в % от годового ИТ-бюджета (сравнение RecSys со средним значением по прочим технологиям)



Продвинутые в ИИ отрасли

Х п. п.

Планируемое изменение ИТ-бюджета RecSys на 2026 г.

	RecSys	Прочие технологии
E-commerce	4,0%	-1,2 п. п. 4,1%
Банкинг и страхование	2,4%	-0,5 п. п. 3,8%
Металлы и горная добыча	2,4%	-0,7 п. п. 3,3%
Строительство и недвижимость	2,2%	-0,8 п. п. 3,2%
Ритейл	2,1%	-0,6 п. п. 2,3%
Нефть и газ	2,0%	0,0 п. п. 3,5%
Медицина и здравоохранение	2,0%	-0,6 п. п. 2,8%
Телеком и медиа	1,8%	+0,2 п. п. 3,8%
ИТ и технологии	1,5%	-0,1 п. п. 3,1%
Электроэнергетика	1,4%	+1,0 п. п. 2,9%
Сельское хозяйство	1,4%	-0,6 п. п. 2,3%
Транспорт и логистика	1,2%	-0,4 п. п. 1,8%
FMCG	0,8%	+0,6 п. п. 1,9%
Машиностроение	0,8%	+0,6 п. п. 1,5%
Автомобильная промышленность	0,8%	0,0 п. п. 1,4%
Химия и нефтехимия	0,7%	+0,4 п. п. 1,4%
Продвинутые в ИИ отрасли	2,4%	-0,4 п. п. 3,7%
Обычные отрасли	1,5%	-0,1 п. п. 2,4%
В среднем	1,7%	-0,2 п. п. 2,7%

Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.3. Какой объем средств в % от годового ИТ-бюджета компания уже инвестировала в технологию за последний год? Вопрос 4.5. Какой объем средств в % от ИТ-бюджета компания планирует инвестировать во внедрение технологии на горизонте года?)

Средний эффект компаний от внедрения рекомендательных систем за последний год, % EBITDA



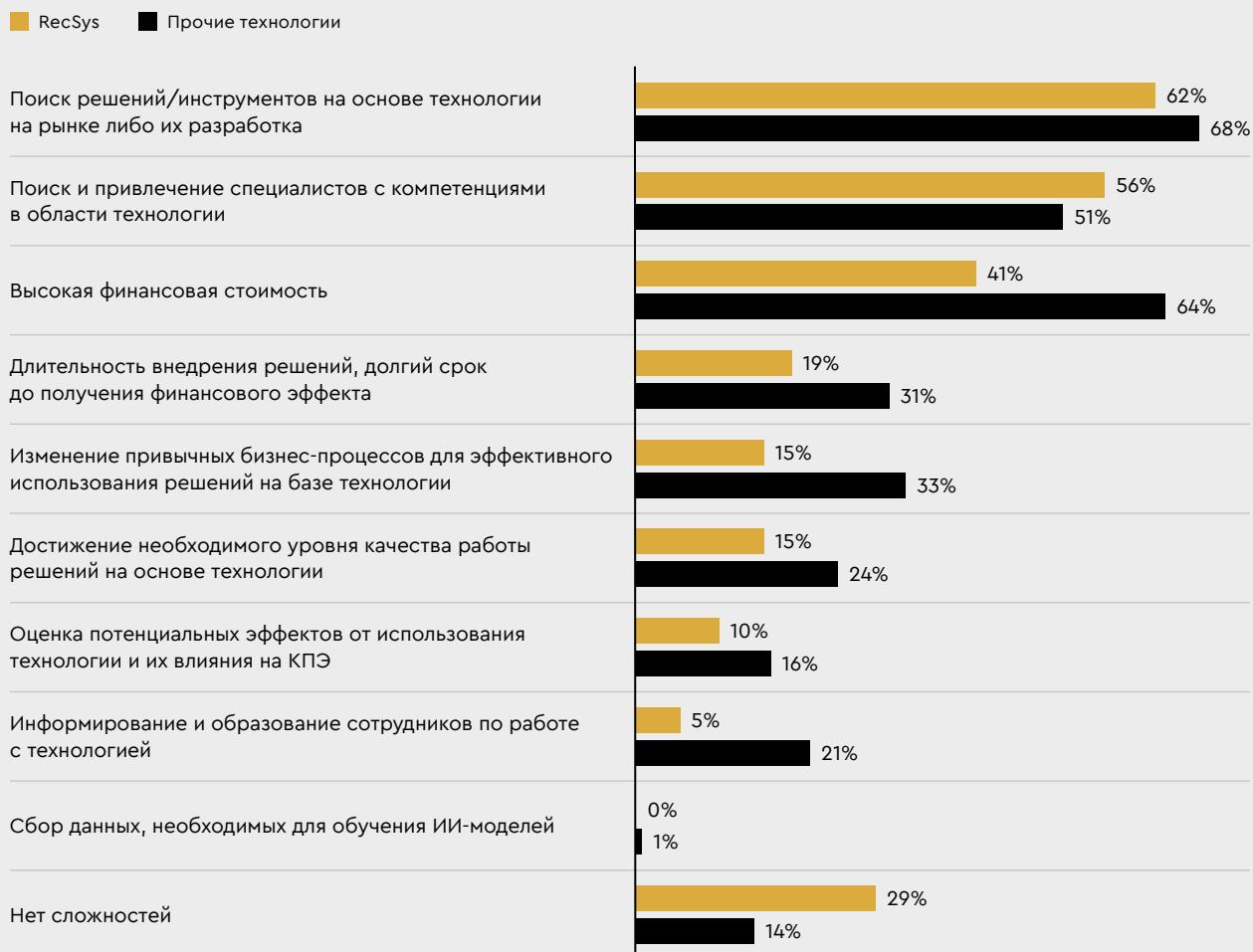
Продвинутые в ИИ отрасли

Банкинг и страхование	2,2%
E-commerce	2,0%
Телеком и медиа	1,9%
ИТ и технологии	1,5%
Металлы и горная добыча	1,2%
FMCG	1,1%
Ритейл	0,8%
Машиностроение	0,8%
Сельское хозяйство	0,8%
Химия и нефтехимия	0,5%
Автомобильная промышленность	0,4%
Электроэнергетика	0,4%
Транспорт и логистика	0,3%
Строительство и недвижимость	0,3%
Медицина и здравоохранение	0,2%
Нефть и газ	0,0%
Продвинутые в ИИ отрасли	1,9%
Обычные отрасли	0,5%
В среднем	0,9%

Источник: опрос индустриальных компаний (блок 4. Эффективность внедрения технологий ИИ, вопрос 4.2. Какой финансовый эффект компания уже получила с момента внедрения технологии в % от годовой EBITDA за последний год?)

Стоит также отметить, что внедрение рекомендательных систем для бизнеса, как правило, является одной из наименьших сложностей для бизнеса: 29% опрошенных компаний отметили, что не сталкивались со сложностями в процессе реализации, против 14% по другим технологиям. При этом 58% опрошенных компаний сообщили, что использовали эти решения три года и более назад, против 32% по другим технологиям. Порог входа для реализации рекомендательных систем ниже и путь к результату короче вследствие невысоких капитальных затрат. Для сравнения: компьютерное зрение требует закупки, установки и обслуживания камер, что привязывает сроки и бюджет к физической инфраструктуре; NLP и речевые технологии зависят от телефонии, стоимости минут распознавания и длительных циклов обеспечения качества, вследствие чего операционные расходы растут, а внедрение замедляется; использование генеративных моделей влечет за собой значимые издержки на инференс и контроль качества ответов.

Сложности, которые видят компании в использовании технологий, % ответов (сравнение RecSys со средним значением по прочим технологиям)



Источник: опрос индустриальных компаний (блок 5. Барьеры и сложности при внедрении технологий, вопрос 5.2. Какие сложности видят компания в использовании технологий?)

37% компаний-респондентов отметили, что Россия находится на уровне стран-лидеров по уровню разработки рекомендательных систем

Отличительная технологическая особенность RecSys — высокие требования к памяти для хранения признаков и эмбеддингов, что может вызывать сложности у ряда компаний. По этой причине вендоры заметно чаще выбирают облачное развертывание: 69% против 56% в среднем по другим технологиям. В целом рекомендательные системы ориентированы на измеримый бизнес-результат и предсказуемую экономику: капитальные затраты минимальны, а масштабирование происходит оперативно. По итогам опроса, 37% компаний считают, что Россия находится на уровне стран — лидеров по разработке рекомендательных систем, тогда как по другим технологиям аналогичный показатель составляет 25–30%.

Оценки пользователей дополняют корпоративную картину. 67% респондентов в целом относятся к рекомендациям положительно (в ИТ-аудитории — 80%), а оценка рисков ниже по сравнению с ИИ-агентами и генеративным ИИ — 9% респондентов считают, что у RecSys больше рисков, чем пользы, против 11 и 14%.

Рынок B2B-решений технологий RecSys

Проведенный опрос показывает, что среди опрашиваемых российских вендоров 11% специализируются на рекомендательных системах. Это значение сопоставимо по масштабу с сегментами генеративного ИИ (около 9%) и речевых/NLP-технологий (около 10%). Однако существуют определенные особенности.

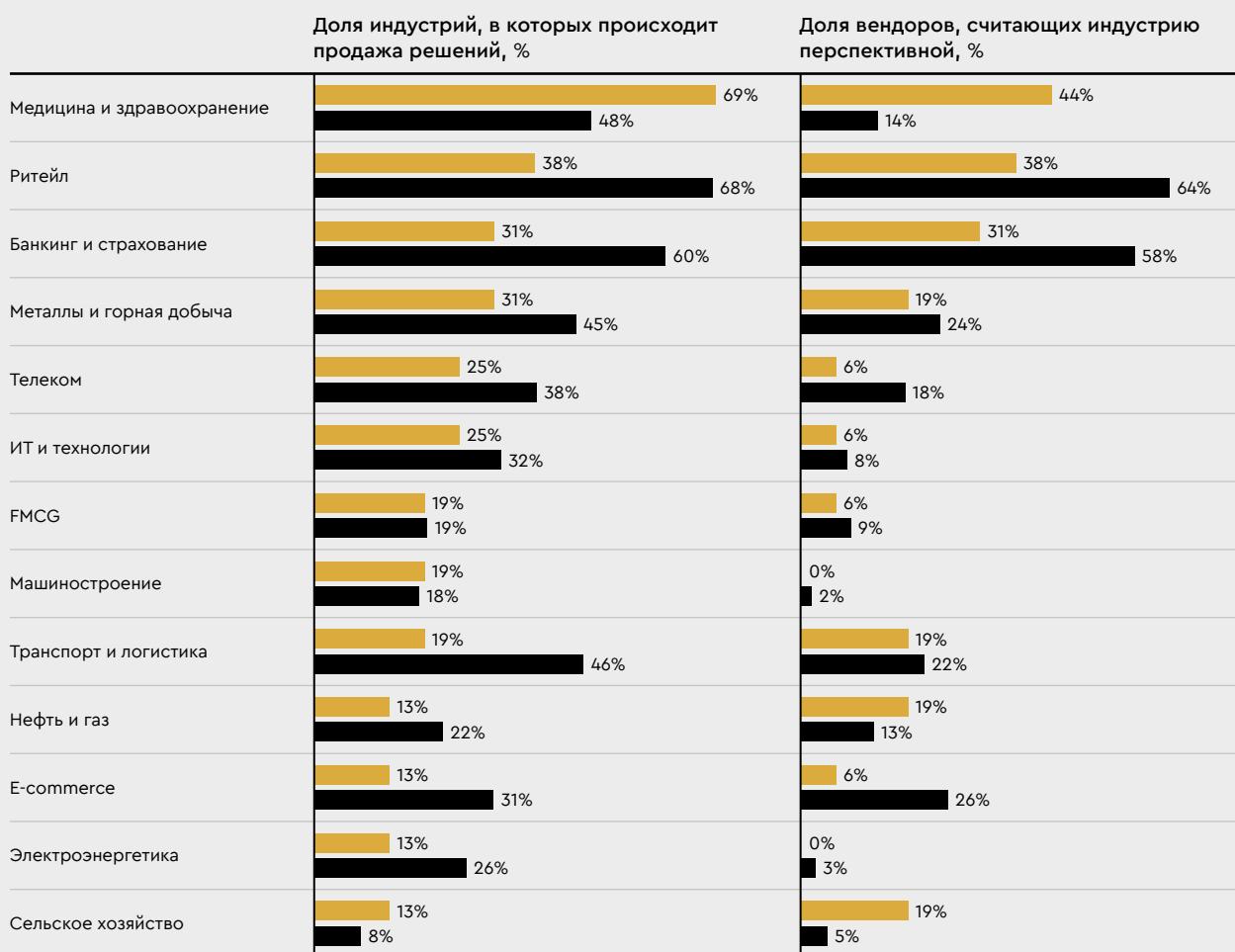
- Более молодой профиль: порядка 25% компаний работают менее трех лет (из них 12% — с опытом до двух лет), тогда как в других технологических сегментах доля игроков со сроком на рынке до трех лет составляет в среднем около 9%.
- Низкая численность персонала: у 58% RecSys-вендоров численность штата ниже 50 человек против 26% у поставщиков иных технологий.
- Низкая выручка: порядка 220 млн руб. на компанию против 677 млн руб. в других сегментах.

В совокупности на рынке вендоров сформировался относительно компактный сегмент по объему выручки, что объясняется простотой внедрения и высоким уровнем развития технологии.

Наибольшая доля клиентов у RecSys-вендоров приходится на медицину и здравоохранение (69% против 48% у прочих технологий), что объясняется дефицитом собственных ИТ-ресурсов в этой отрасли и потребностью во внешних компетенциях при соблюдении требований безопасности. В e-commerce наблюдается обратная картина: доля клиентов в сравнении с прочими технологиями ниже (13% против 31%), как и доля вендоров, считающих индустрию перспективной для создания внешних решений (6% против 26%), поскольку e-commerce выступила пионером внедрения, имеет высокий уровень развития и уже не нуждается в услугах вендоров. В сельском хозяйстве, напротив, доля клиентов и ожидания роста выше среднерыночных для RecSys (13% против 8% по текущим продажам и 19% против 5% по перспективам), что говорит о низком развитии рекомендательных систем в этой индустрии.

Индустрии, в которых происходят продажи у вендоров/доля вендоров, считающих индустрию перспективной, % ответов (сравнение RecSys и среднего значения по прочим технологиям)

■ RecSys ■ Прочие технологии



Источник: опрос вендоров (блок 4. Портфель клиентов, вопрос 4.1. В каких индустриях компания продает решения на основе технологии и какие считает наиболее перспективными?)

Будущее и перспективы развития технологий рекомендательных систем

На фоне описанной структуры спроса и инвестиций рекомендательные системы выступают одной из наиболее зрелых технологий в выборке. Этот вывод подтверждают и компании-пользователи, и вендоры: 62% опрошенных компаний и 80% вендоров считают, что основная часть потенциала RecSys уже реализована; еще 24 и 20% соответственно отмечают полную реализованность потенциала. Ожидания по масштабу применения также стабильны: 56% компаний (против ~40% по другим технологиям) и 31% вендоров (против ~19%) полагают, что технология продолжит использоваться в том же масштабе, что и сейчас.

На сегодняшний день технология уже достаточно распространена, а ее дальнейшее развитие не столь стремительно ввиду определенных барьеров к развитию:

- Слабая и косвенная обратная связь. В большинстве сервисов обратная связь ограничена косвенными действиями (просмотр, клик), которые не всегда отражают истинные предпочтения пользователя. Это затрудняет обучение моделей и снижает точность рекомендаций.
- Ограничения памяти и производительности. Для того чтобы укладываться в требования по скорости и ресурсам, онлайн-модели часто работают с коротким контекстом — видят только недавние действия пользователя. Это ухудшает персонализацию в сложных сценариях. Попытка хранить и подавать длинную и разнообразную историю взаимодействий для миллионов пользователей резко увеличивает нагрузку на память, сеть и инфраструктуру с низкими задержками, что усложняет масштабирование.
- Неравный доступ к данным для исследований. Крупные массивы поведенческих логов и контента сосредоточены у больших платформ, что ограничивает возможности для воспроизводимых экспериментов и замедляет открытие сравнения подходов для академического сообщества и малого бизнеса.

Рекомендательные системы — уже зрелая технология, однако их развитие может получить дополнительный импульс благодаря интеграции с генеративными моделями

С другой стороны, сохраняется потенциал постепенного роста за счет интеграции с другими технологиями. Как отмечалось ранее, рекомендательные системы уже сейчас интегрируются с генеративными моделями и такая связка может улучшить качество в ряде сценариев (уточнение намерений, объяснимость, персональные коммуникации) рекомендательных систем за счет следующих факторов:

- Повышение качества рекомендаций. Генеративная модель уточняет намерение пользователя и обогащает признаки объектов, за счет чего входные данные для скоринга становятся информативнее. В результате система точнее оценивает вероятности целевых действий и параметры показа — релевантность, уместность, новизну, соблюдение правил и ограничений — и лучше калибрует ранжирование и пересортировку списка. Это повышает качество и конверсию при неизменном трафике и упрощает запуск новых позиций без накопленной истории.
- Усиление объяснимости. Генеративная модель формирует понятные пользователю пояснения на основе фактических сигналов (история действий, атрибуты объекта, действующие ограничения), выводит ключевые причины показа и альтернативы, а также фиксирует обоснование. На практике это размещается в карточке товара, письмах и чат-боте, снижает отказы и повышает доверие к выдаче.
- Расширение персональных коммуникаций. Модель автоматически подбирает тон, формат, канал и время сообщения под сегмент и текущий контекст, генерирует варианты заголовков, текстов и кратких аннотаций для электронных писем, push-сообщений и баннеров, а затем отбирает лучшие. Это ускоряет подготовку кампаний и повышает отклик без увеличения частоты контактов.

Таким образом, дальнейшее развитие генеративных моделей может положительно влиять на развитие рекомендательных систем, обогащая их функционал.

Заключение

Хотя интерес к искусственному интеллекту (ИИ) приобрел массовый характер сравнительно недавно, многие его направления уже достигли стадии технологической зрелости. Компьютерное зрение (CV), рекомендательные системы (RecSys), обработка естественного языка (NLP) и речевые технологии демонстрируют устойчивое масштабирование, стабильный рост бизнес-эффекта и расширение прикладных сценариев. Для этих технологий характерна эволюционная, а не революционная динамика развития: повышение устойчивости и надежности, снижение совокупной стоимости владения (в том числе за счет оптимизации инференса), улучшение интеграции в ИТ-инфраструктуру и повышение удобства эксплуатации. Их значение для эффективности бизнес-процессов и роста выручки остается системообразующим. Ключевой запрос пользователей — прикладная полезность: важнее, чтобы нейросети помогали в повседневных задачах и были встроены в привычные сервисы, чем демонстрировали успехи на абстрактных бенчмарках. Именно продуктовая интеграция ИИ в существующие пользовательские сценарии является драйвером развития рынка.

Традиционный подход к классификации технологий искусственного интеллекта предполагал деление на четыре независимые области, каждая из которых была покрыта в исследовании. Однако анализ текущих трендов и данных настоящего исследования позволяет переосмыслить роль генеративного ИИ: это не четвертая вертикальная технология, стоящая в одном ряду с остальными, а горизонтальная платформа, которая фундаментально трансформирует все три классические области изнутри. При этом в ближайшие годы основные структурные изменения будут приходить на связку foundation-моделей и агентских систем.

Современные модели перестают быть пассивными хранилищами знаний и становятся активными исполнителями: извлекают, проверяют и синтезируют информацию, а затем инициируют действия через подключение к внешним базам, сервисам и API. Отдельным слоем поверх моделей формируется агентская архитектура: ИИ-агенты планируют шаги, выбирают инструменты, обращаются к внешним сервисам и координируются между собой. Корректнее рассматривать их как операционный слой исполнения, а не как отдельную технологию.

Генеративный ИИ — не четвертая вертикаль, а горизонтальная платформа, переосмысливающая NLP, CV и RecSys. Рынок входит в эпоху foundation-моделей и агентских систем: модели становятся активными исполнителями, способными решать множество задач в единой архитектуре

Суть сдвига заключается в переходе от специализированных узкопрофильных моделей к универсальным foundation-моделям. Речь идет не об одной супермодели, а о платформе, где задачи решаются композицией нескольких специализированных модулей в единой архитектуре. В старой парадигме каждая задача требовала отдельной модели: чат-бот на русском языке — одна модель, анализ рентгеновских снимков — другая, персонализированные рекомендации — третья. Теперь единая платформа объединяет навыки и подключает инструменты, а множество задач решается через оркестрацию моделей внутри общего контура.

Фактически ИИ становится операционной системой нового цифрового мира, аналогом электричества для технологической среды. Уже формируются протоколы взаимодействия между ИИ-ассистентами, а значит, архитектура цифрового потребления переходит от работы с приложениями к взаимодействию с интеллектуальными сервисами, действующими от имени пользователя. При этом развитие ИИ-агентов, способных самостоятельно выполнять многошаговые задачи и закрывать действия за человека, ускоряется, и по этому направлению Россия не отстает: идут пилоты и развертывания в различных индустриях.

При этом ключевые барьеры индустрии в целом сместились с уровня аппаратных ограничений на уровень экономики и социальной готовности к внедрению технологий. Основными ограничителями становятся стоимость масштабирования (инфраструктура, инференс, перестройка процессов) и консерватизм пользователей и корпораций. Решением становятся витринные кейсы с измеримым эффектом, стандартизованные метрики и упрощенный, прозрачный путь передачи данных, обеспечивающий доверие и выгоду для конечного пользователя.

Российская экосистема ИИ на этом фоне демонстрирует ускоренный переход от стадии интереса к стадии системного внедрения: в перспективе 3–5 лет главный вызов ИИ не технологический, а внедренческий. По данным опросов технических директоров и вендоров, пилоты по всем технологиям запущены у порядка 80% крупных компаний, а доля фактического применения генеративного ИИ достигла 71% в 2025 г. (+17 п. п. год к году). Однако сохраняется характерный разрыв между уровнем инженерных компетенций и скоростью интеграции решений в операционные процессы — в первую очередь из-за консервативности ряда индустрий, настороженности к обработке данных (лишь 50% респондентов пользовательского опроса считают, что пользы от ИИ больше, 42% — что польза и риски сопоставимы, 8% — что рисков больше) и дефицита показательных кейсов с конкретным результатом. Исторически подобные барьеры преодолеваются через массовые сервисы, нормализующие взаимодействие с технологией (аналогично тому, как интернет-сайты и социальные сети стали повседневными), либо экзогенные события, радикально повышающие полезность технологий (пример — бум сервисов доставки и видеоконференций во время пандемии COVID-19). В текущем цикле такими катализаторами станут сервисы на базе генеративного ИИ и ИИ-агентов с понятным бизнес-эффектом и ценностью для пользователей.

Несмотря на это, экономический эффект от ИИ уже виден. В продвинутых отраслях — e-commerce, ИТ, телекоме и медиа — его вклад достигает 8% EBITDA, а ожидаемый совокупный эффект ИИ на EBITDA на горизонте ближайших 12 месяцев оценивается в 13–21% при среднем по рынку уровне 11%. По оценкам, к 2030 г. годовой эффект от внедрения ИИ в российской экономике составит 7,9–12,8 трлн руб. Это эквивалентно величине до 5,5% ВВП, сопоставимой с прибылью всей банковской отрасли. При этом драйвером эффекта выступает не только сокращение издержек (на это указывают 90% компаний), но и рост выручки за счет персонализации сервисов, ускорения вывода новых продуктов и повышения вовлеченности клиентов, особенно в сегментах пользователей. Важно понимать, что на рынке уже присутствуют отечественные решения, конкурирующие с глобальными лидерами (в том числе уровня DeepSeek и ChatGPT) в отдельных задачах и сегментах.

Внедрение ИИ по отраслям также имеет свою динамику. Локомотивами остаются телеком и медиа, e-commerce, банкинг и страхование, ИТ и технологии благодаря высокой цифровой зрелости, гибкой инфраструктуре и культуре экспериментов. В прочих отраслях, например в производстве и добыче, максимальный эффект дают CV-кейсы контроля качества, безопасности и предиктивного обслуживания. В медицине потенциал применения высок, но его реализация зависит от регуляторики, стандартизации данных и готовности к процессным изменениям со стороны индустрии.

Организационные модели компаний также эволюционируют. Если раньше ИИ-проекты запускались как пилоты в отдельных подразделениях, то сегодня они переходят к стадии устойчивой операционализации. В 90% крупных компаний ответственность за развитие ИИ закреплена на уровне CEO-1, а четверть таких компаний уже имеют выделенные ИИ-функции. Наиболее эффективной оказывается гибридная модель — центр экспертизы в сочетании со встроенными ИИ-ролями в бизнес-подразделениях. Это позволяет ускорить внедрение за счет использования готовых компонентов, адаптации open-source- и коммерческих моделей и создания партнерских экосистем с технологическими корпорациями и нишевыми вендорами.

На макроуровне Россия имеет высокий уровень в технологическом развитии и общем внедрении ИИ, несмотря на не самую высокую развитость науки в рамках технологии и низкую доступность аппаратной базы. Это формирует прагматичный профиль развития: использование open-source- и партнерских экосистем, развитие суверенной инфраструктуры там, где это оправдано по совокупной стоимости владения и рискам, а также концентрация усилий на формировании и удержании талантов за счет создания комфортной среды для их развития. Вклад отдельных специалистов в итоговый результат остается непропорционально высоким, что делает кадровую политику одним из стратегических факторов конкурентоспособности на рынке.

Таким образом, зрелые технологии ИИ — CV, RecSys, NLP и речевые системы — продолжают повышать эффективность и обеспечивают ресурсную базу для перехода к следующему этапу — сервисам поверх моделей. При этом генеративный ИИ будет активно развиваться как горизонтальная трансформация, параллельно охватывающая каждую из технологий. Данные исследования подтверждают этот тренд: раздел NLP прямо указывает на «замену классических подходов на генеративные модели». Так, 70% компаний внедрили NLP-решения в последние два года, причем значительная часть — на базе генеративного ИИ. Компьютерное зрение активно интегрируется в мультимодальные генеративные модели, а рекомендательные системы обогащаются возможностями LLM для создания объяснимых рекомендаций. Как отмечалось в основной части исследования, «среди ключевых технологий на рынке ИИ потенциальный эффект в будущем обратно пропорционален их текущей зрелости — генеративный ИИ и компьютерное зрение возглавляют ИИ-трансформацию». Таким образом, правильнее говорить не о четырех параллельных технологиях, а о трех классических областях применения (NLP, CV, RecSys), каждая из которых переживает трансформацию под воздействием генеративного ИИ как горизонтальной технологической платформы.

К 2030 г. выиграют те компании, которые уже сейчас вводят понятные общие стандарты качества ИИ-сервисов, встраивают ИИ в ключевые процессы и выстраивают устойчивую работу с сильными специалистами. Масштаб проникновения ИИ в экономику и общество становится вопросом управляемости изменений и зрелости экосистемы внедрения, а распределение экономической ценности — вопросом качества сервисов, стандартов и организационных решений. Именно способность превратить ИИ в удобные повседневные сервисы и агентские действия для пользователя станет главным источником устойчивого преимущества.

Историческая параллель очевидна: как когда-то интернет сделал структурный сдвиг, так в ближайшие годы то же сделает искусственный интеллект. Следующий цикл — внедрение. Чтобы этот эффект стал массовым и измеримым, важно быстро запускать сервисы и накапливать рабочие кейсы применения.

Опираясь на результаты исследования, мы, «Яков и Партнёры» и «Яндекс», видим для всех рассматриваемых технологий следующий набор приоритетных действий для ключевых стейкхолдеров:

Для компаний

- Построение LLM-платформ как единого контура доступа к моделям, данным, инструментам и метрикам качества. При возможности — облачное развертывание: оно обеспечивает эластичность, ускоряет эксперименты и дает более прогнозируемую экономику владения, позволяя запускать концепты и масштабироваться без существенных капитальных затрат. Конкретный выбор провайдера и архитектуры определяется требованиями безопасности, приватности и SLA.
- Держать фокус на нескольких приоритетных кейсах интеграции ИИ в функции для четкого анализа эффекта, закладывая реалистичные сроки достижения полномасштабного эффекта (от 12 до 24 месяцев), без поиска быстрых побед.
- Регулярно проводить собственный анализ внешних бенчмарков через обмен практиками с коллегами в РФ и в мире.
- Отдельный приоритет — встраивание ИИ в привычные пользовательские сервисы и сценарии, где прикладная выгода очевидна.
- Сфокусировать усилия на отечественных решениях, стимулируя отрасль, и сопровождать их развитие открытой конструктивной критикой, чтобы российские компании ускоряли прогресс и уверенно конкурировали на мировом рынке.

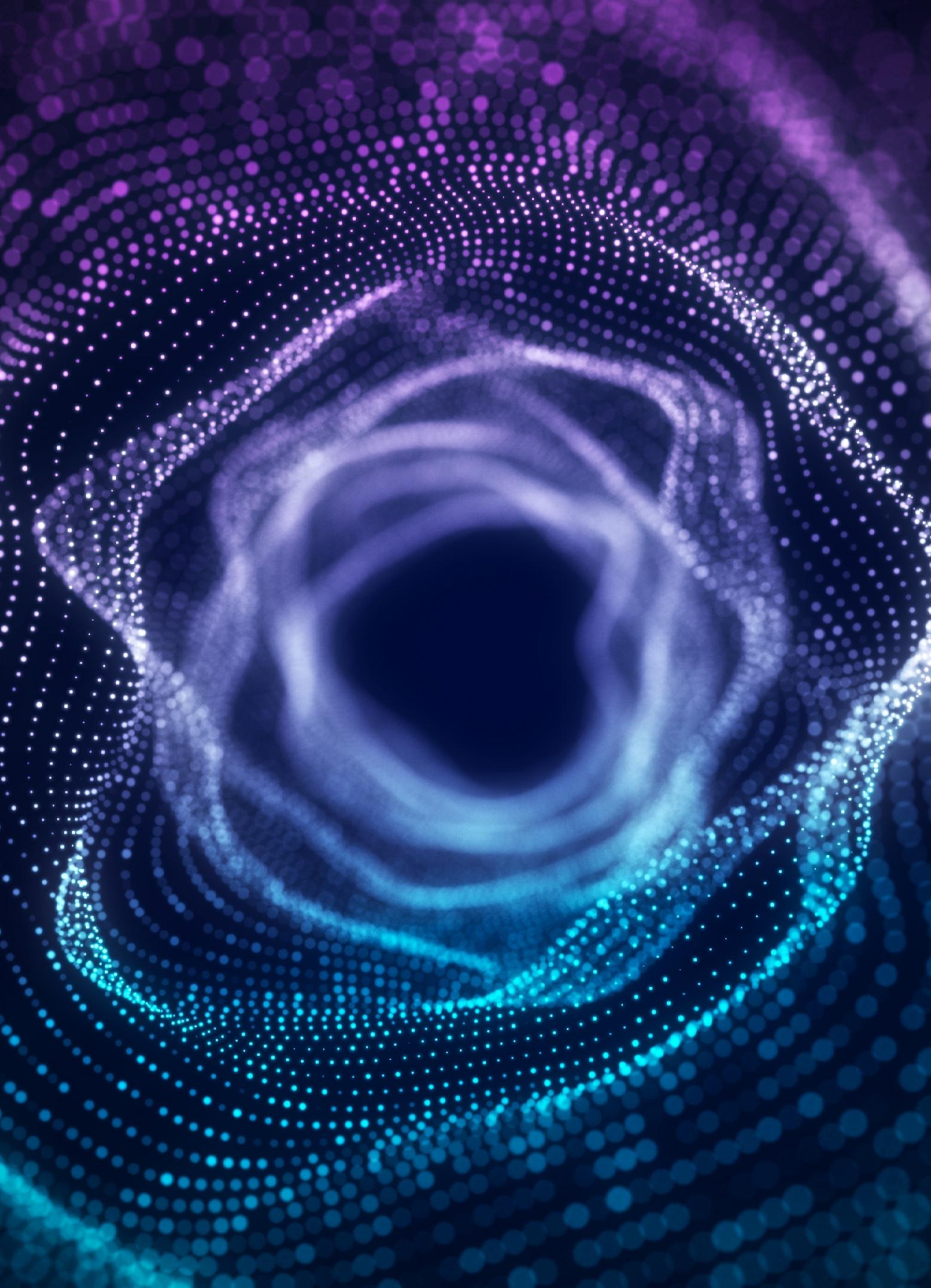
Для государства

- Расширить финансовые и нефинансовые меры поддержки отраслей и компаний: предоставлять гранты и субсидии на НИОКР и внедрение, вводить налоговые стимулы, обеспечивать льготное финансирование, а также системно продвигать повестку практической ценности ИИ для населения и бизнеса. Эти меры направлены на снижение совокупных издержек разработки и опытной эксплуатации решений, а также на ускорение вывода технологий в приоритетные сферы — медицину, образование и социальный сектор.
- Внедрять умное, рискоориентированное регулирование, взвешивая все негативные и положительные факторы, чтобы не замедлять развитие технологии.

Для людей

- Системно осваивать ИИ-инструменты в повседневной деятельности — от обучения до автоматизации рутинных задач, в том числе развивать ключевые компетенции (работа с данными, критическая оценка результатов) и преодолевать естественный скепсис через практику. Как с переходом на смартфоны: только собственный опыт применения позволяет раскрыть потенциал технологий и не упустить рыночные преимущества.
- Осознанно управлять персональными данными: понимать объем и цели передачи, давать обратную связь по качеству пилотов и сервисов в качестве механизма улучшения результатов.

Только при согласованных действиях бизнеса, государства и академического сообщества — при активном участии самих пользователей — возможно масштабное, безопасное и экономически обоснованное внедрение ИИ на благо людей и страны.



Примечания

1. Прогноз Минэкономразвития.
2. Данные Mediascope за ноябрь 2025 г.
3. IBM: <https://www.ibm.com/think/topics/computer-vision>
4. <https://dl.acm.org/doi/10.1145/138859.138867>
5. <https://dl.acm.org/doi/10.5555/2074094.2074100>
6. <https://googleblog.blogspot.com/2005/06/search-gets-personal.html#:~:text=With%20the%20launch%20of%20Personalized,tell%20us%20what%20you%20think>
7. Huang J. et al. A comprehensive survey on retrieval methods in recommender systems//ACM Transactions on Information Systems. – 2024. <https://dl.acm.org/doi/pdf/10.1145/3771925>
8. https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm

Вся информация, содержащаяся в настоящем документе (далее также «Исследование», «Материалы Исследования»), предназначена только для информационных частных некоммерческих целей и не является профессиональной консультацией или рекомендацией. Ни информация, содержащаяся в Исследовании, ни ее использование любым лицом не создают договора, соглашения или отношений между компанией «Яков и Партнёры» и любым лицом, получившим и рассматривающим Материалы Исследования и (или) любую информацию, содержащуюся в Исследовании.

«Яков и Партнёры» оставляют за собой право вносить изменения в информацию, содержащуюся в Исследовании, однако не берут на себя обязательств по обновлению такой информации после даты, указанной в настоящем документе, несмотря на то что информация может стать устаревшей, неточной или неполной. «Яков и Партнёры» не дают обещаний или гарантий относительно точности, полноты, адекватности, своевременности или актуальности информации, содержащейся в Исследовании.

«Яков и Партнёры» не проводили независимую проверку данных и предположений, использованных в Исследовании. Изменения в исходных данных или предположениях могут повлиять на анализ и выводы, представленные в Исследовании. «Яков и Партнёры» не предоставляют юридических, нормативных, бухгалтерских, финансовых, налоговых, регуляторных консультаций. Любое лицо, получившее и рассматривающее Материалы Исследования и (или) любую информацию, содержащуюся в Исследовании, несет ответственность за получение независимой консультации в вышеуказанных областях.

Консультации в вышеуказанных областях могут повлиять на анализ и выводы, представленные в Исследовании. Ничто в Исследовании не подразумевает рекомендаций о совершении действий, которые могут приводить к нарушению любого применимого законодательства. «Яков и Партнёры» не предоставляют заключений о справедливости рыночных сделок или оценок таких сделок. На Материалы Исследования нельзя полагаться как на такие заключения или оценки, и их не следует толковать как таковые. Материалы Исследования могут содержать прогнозные данные (включая рыночные, финансовые, статистические данные, но не ограничиваясь ими), будущая реализация которых не является гарантированной. Вследствие этого такие прогнозные данные связаны с некоторым труднопредсказуемым риском и неопределенностью. Фактические будущие результаты и тенденции могут существенно отличаться от описанных в прогнозах вследствие целого ряда разных факторов. Если какое-либо лицо полагается на информацию, содержащуюся в Материалах Исследования, то оно делает это исключительно на свой собственный риск. Никакие гарантированные имущественные права не могут быть получены из любого вида информации, представленной в Исследовании. В максимальной степени, разрешенной законом (и за исключением случаев, когда иное согласовано с «Яков и Партнёры» в письменной форме), «Яков и Партнёры» не несут никакой ответственности за любой ущерб, который может быть причинен в любой форме любому лицу вследствие использования, неполноты, некорректности, неактуальности любой информации, содержащейся в Исследовании. Материалы Исследования ни полностью, ни частично нельзя распространять, копировать или передавать какому-либо лицу без предварительного письменного согласия «Яков и Партнёры». Материалы Исследования являются неполными без сопроводительного комментария, и на них нельзя полагаться как на отдельный документ. Любое лицо, получившее и рассматривающее Материалы Исследования и (или) любую информацию, содержащуюся в Исследовании, настоящим отказывается от любых прав и требований, которые оно может иметь в любое время против «Яков и Партнёры» в отношении Исследования, содержащейся в Исследовании информации или других связанных с Исследованием материалов, выводов, рекомендаций, включая их точность и полноту. Названия продуктов, логотипы и товарные знаки компаний, указанные в настоящем документе, охраняются законом. Получение и рассмотрение настоящего документа считается согласием со всем вышеизложенным.

Искусственный интеллект в России — 2025: тренды и перспективы

Контент и аналитика отчета подготовлены консалтинговой компанией «Яков и Партнёры»:

Максим Болотских, партнер, руководитель практики ИИ и высоких технологий

Марина Дорохова, директор, практика ИИ и высоких технологий

Иван Серов, консультант

Максим Шуркин, консультант

Сабина Бабаева, старший бизнес-аналитик

Даниил Ермолаев, бизнес-аналитик

Команда «Яков и Партнёры», выпустившая материал:

Никита Драль, дизайнер

Дарья Борисова, дизайнер

Ксения Чемоданова, выпускающий редактор

Команда «Яндекса»:

Артур Самигуллин, руководитель платформы Yandex AI Studio в Yandex B2B Tech

Елена Белоброва, руководитель направления по развитию ML-сервисов в Yandex B2B Tech

Николай Савушкин, руководитель службы рекомендательных технологий

Сергей Овчаренко, руководитель отдела мультимодальных анализа и генерации

Алексей Колесов, СТО R&D

«Яков и Партнёры» — международная консалтинговая компания со штаб-квартирой в Москве и обширной сетью экспертов и партнеров более чем в 120 странах мира. Мы увлечены задачами социального и экономического развития и работаем плечом к плечу с лидерами различных отраслей экономики и общественного сектора, вместе формируя поворотные моменты в истории компаний и общества, добиваясь устойчивых результатов, масштабы которых выходят далеко за рамки отдельных бизнесов.

© «Яков и Партнёры», 2025. Все права защищены.

Связаться с авторами, запросить комментарии, а также уточнить ограничения по использованию и перепечатке материалов можно направив запрос на адрес:

media@yakovpartners.ru

Больше исследований, аналитики и публикаций — на сайте:

www.yakovpartners.ru



Яков и Партнёры × Яндекс

© ООО «Яков и Партнёры», 2025
Все права защищены

www.yakovpartners.ru

-  [YakovPartners](#)
-  [yakov.partners](#)
-  [yakov-partners](#)