

Содержание

Клиентский генеративный ИИ в финансовых организациях: что показали полевые интервью

Обезличенная сводка серии экспертных интервью (2026).

Названия и размер организаций, страны и юрисдикции не раскрываются намеренно. Текст переписан и обобщён: совпадение формулировок с исходными материалами исключено, сохранены только факты и выводы.

О чём материал: как финансовые организации внедряют генеративный ИИ в клиентские каналы — чат-боты, ассистенты в приложении, голосовые боты, которые общаются с клиентом напрямую. Уровень внедрения, выбор технологий, сценарии, проблемы, регуляторика и контроль, кто принимает решения, планы по сегментам.

Условные обозначения участников (нужны, чтобы разделы ссылались друг на друга):

- **Банк А** — крупный; покупает фронтирные модели, прикладной слой строит сам;
- **Банк Б** — крупный; покупает готовые решения у вендоров;
- **Банк В** — крупный; больше двух лет безуспешно переводит legacy-бота на GenAI;
- **Группа Г** — крупная банковская группа; около двух лет сознательно не выпускала ИИ к клиентам (информация через консультанта);
- **Банк Д** — цифровой; самый зрелый кейс выборки, ассистент — основной канал;
- **Банк Е** — средний; open-source модели на своём железе внутри контура; действует постоянный механизм компенсаций клиентам за ошибки бота;
- **Банк Ж** — ещё не внедрил; идёт формальный тендер; голосовой контакт-центр; данные строго локально;
- **Банк З** — региональный; собирает агентские связки на low-code;
- **Платёжный провайдер; B2B-кредитная организация; Платёжная B2B-компания;**
- **Вендор диалогового ИИ; Технологическая дочка банка;**
- **Консультанты** по AI-governance, комплаенсу и аудиту.

Ключевые цифры: 16 бесед · 17 участников · 5 типов регуляторных режимов · 12 сценариев × 5 сегментов · 7 классов проблем · лучший показатель решения диалога без эскалации — 88-89% · прогноз вендора по массовому голосу — 1-1,5 года.

1. Выборка и почему ей хватает веса

За примерно два месяца провели 16 разговоров с 17 участниками. Пятнадцать — полноценные структурированные интервью по 30-40 минут (видео или телефон, формат discovery: открытые вопросы о практике респондента, без продажи продукта); один ранний контакт ограничился обменом материалами и в анализ не вошёл. Респондентов нашли тремя путями: тёплые профессиональные связи, профессиональные и alumni-сообщества, платный скрининг через экспертную сеть — последний канал важен тем, что достаёт практиков «в роли» из крупных организаций, до которых холодный outreach не доходит.

Кто в выборке:

Сегмент	Интервью	Роли
Крупный банк	2 напрямую (+1 исключённый короткий контакт; +2 группы косвенно)	Комплаенс второй линии; старший дата-саентист по модельному риску
Цифровой банк	3	Архитектор ИИ-инфраструктуры; руководитель команды чат-бота; руководитель клиентского сервиса
Финансовая B2B-компания	3	Комплаенс-офицер, он же операционный владелец работающего клиентского ИИ; ИИ-инженер (внутренние агенты); владелец
Вендор	2	Основатель/СЕО платформы диалогового ИИ; руководитель ИИ в технологической дочке банка (на том

Сегмент	Интервью	Роли
Консультант по ИИ и комплаенсу	5	же звонке был инженер второго регионального банка) CRO крупной банковской группы; руководитель ИИ в инвесткомпании (раньше — практик ИИ в банке); соуправляющий консалтинга по комплаенсу и моделированию рисков; аудитор по регуляторному комплаенсу; независимый консультант по AI-governance

Выборка старшая по составу: трое основателей/владельцев, один бывший C-level (CRO крупной группы), трое руководителей функций, семь старших практиков (архитектура и инженерия ИИ, модельный риск, комплаенс второй линии — включая человека, лично спроектировавшего процесс одобрения ИИ в одном из самых продвинутых цифровых банков), двое старших консультантов/аудиторов. Функционально покрыты все три линии защиты, инженерия, сторона предложения (вендоры) и консультанты, чьи клиентские портфели дают видимость по десяткам организаций сразу.

Почему на такой выборке можно строить выводы. Во-первых, покрыт весь спектр зрелости — от банка, который ещё ничего не внедрил (тендер в процессе), до цифрового банка с одним из самых развитых ИИ-хозяйств в отрасли (сотни оценённых по риску LLM-кейсов в проде, живой голосовой ассистент). Во-вторых, покрыты контрастные регуляторные режимы — зрелые надзорные, надзор за модельным риском, переходные и режимы с абсолютной локализацией данных; регуляторика становится наблюдаемой переменной, а не константой. В-третьих, есть и спрос, и предложение, и советники, видящие много организаций одновременно, — самоотчётность удаётся перекрёстно проверять.

Ограничения, которые стоит держать в голове: один респондент на организацию (кроме одного звонка с двумя организациями); перекося

сторону практиков, а не держателей бюджета; все данные — самоотчётность. Там, где вывод опирается на один голос, это оговорено.

2. Текущее состояние: GenAI перед клиентом

Общая картина

Каждая розничная организация выборки либо уже эксплуатирует клиентского чат-бота, либо его закупает — но генеративный ИИ перед клиентом пока исключение, а не норма. Разброс широкий: основной трафик по-прежнему несут legacy-боты на скриптах и NLU; GenAI в проде — у цифровых игроков и части крупных incumbent'ов; голос работает ровно у одной организации выборки. Зрелость коррелирует с сегментом и операционной культурой, а не с размером рынка: средний банк (Е) заметно впереди крупного incumbent'а, трижды встречающегося в материалах, а лидирующий цифровой банк (Д) по собственной правдоподобной оценке опережает общий рынок примерно на четыре года.

Второе сквозное наблюдение: у консервативных организаций внутренний ИИ на годы опережает клиентский. Группа Г около двух лет намеренно не пускала диалоговый ИИ к клиентам — по словам представителя группы, она принципиально не была готова передать ИИ контроль над клиентской коммуникацией, — параллельно активно внедряя ИИ в разработку, почту и исследования. B2B-кредитная организация эксплуатирует сложную внутреннюю мультиагентную аналитику примерно для сотни аналитиков и вообще не планирует клиентский ИИ. Там, где цифровые игроки ведут клиентским каналом, incumbent'ы ведут бэк-офисом.

Зрелость по сегментам

Сегмент	Стадия клиентского GenAI	Каналы	Голос	Ориентировочные объёмы
Крупный банк	Широкий разброс внутри сегмента: от прода 2–5 лет (А, Б) до стагнации (В: 2+ года безуспешно й конверсии NLU→GenAI) и	Мобильное приложение , веб; копилоты для операторов	Есть, но ограничено (Б: используетс я меньше чата и копилотов); elsewhere — нет или не обсуждалось	Миллионы клиентов и тысячи диалогов в день (А); тысячи–десятки тысяч взаимодействий в день (Б)

Сегмент	Стадия клиентского GenAI	Каналы	Голос	Ориентировочные объёмы
	сознательно го сдерживания (~2 года, группа Г)			
Цифровой банк	Весь спектр зрелости: самое развитое продакшн-хозяйство выборки (Д: ассистент как ядро канала) · боевой RAG-бот рядом с выводимым кнопочным ботом (Е) · до внедрения, тендер (Ж)	Ассистент в приложении и текстовый чат; у Ж — голосовой контакт-центр (сотни операторов), текст маргинален	Живой (Д) · планируется (Е) · приоритет №1, запуск в горизонте ~2 лет (Ж)	Сотни LLM-кейсов, 100+ действий ассистента, ~88-89% диалогов без эскалации (Д) · >15% всего клиентского трафика (Е) · ноль при десятках тысяч звонков в день у людей (Ж)
Финансовая B2B-компания	От раннего прода/экспериментов (платёжный провайдер сам оценивает автоматизацию как стартовую точку) до полного отсутствия клиентского ИИ (только внутренние агенты; живой саппорт как	Приоритет тексту; живой чат в интернет-банке там, где ИИ нет	Экспериментальный, ограничен пресетами	~5-10% саппорта автоматизировано (провайдер); ноль у остальных

Сегмент	Стадия клиентского GenAI	Каналы	Голос	Ориентировочные объёмы
Вендор (взгляд на клиентский рынок)	премиум) Рынок ранний: один крупный банк вендор характеризует как сильно отстающий (шестилетний цикл внедрения); самый продвинутый банк его домашнего рынка только вывел голос в полный прод	Сначала чат	Массовый голос — через 1-1,5 года	—
Консультанты (взгляд на клиентский рынок)	Клиенты с активами до ~\$5 млрд ещё не внедряют (вопрос времени, не принципа); внедрившиеся — выше этого порога; теневое использование ИИ повсеместно при официально отрицании	—	—	—

Что означает разброс по каналам и объёмам

Текстовый чат — универсальный первый канал (приложение и веб); мессенджеры маргинальны, а у Ж им прямо не доверяют в банковских вопросах. Голос — фронтир: живой у одного цифрового банка (Д), вторичный у одного крупного банка (Б), экспериментальный и скованный пресетами у платёжного провайдера и заявленный приоритет №1 (но реальность горизонта ~2 лет) у Ж — там контакт-центр голосовой, и автоматизация голоса ценна именно мгновенной разгрузкой операторов. Оценка вендора — сначала чат, массовый голос через 1-1,5 года, текущий встроенный голос ещё с высокой задержкой — совпадает с тем, что видно по организациям.

Объёмы охватывают четыре порядка: от нуля (до внедрения) до сотен LLM-кейсов и тысяч диалогов в день. Две организации назвали долю автоматизации: >15% всего клиентского трафика (Е) и 5-10% саппорта (провайдер, по собственной оценке — ранняя стадия). Самый зрелый игрок измеряет ~88-89% решения без эскалации к человеку — ориентир того, чего достигает зрелый GenAI-саппорт.

3. Технологии, «строить или покупать», требования к поставщикам

Как выглядят стеки

На зрелом полюсе выборки воспроизводится один и тот же архитектурный паттерн: фронтирные коммерческие LLM за собственным слоем оркестрации и контроля.

- Банк А покупает сторонние модели — несколько семейств фронтирных коммерческих LLM — через облачный ИИ-маркетплейс с версионированием и откатом, а прикладную нагрузку строит внутри. Обоснование: коммерческие модели превосходят всё, что банк способен разработать сам.
- Платёжный провайдер строит собственный программный слой поверх коммерческих LLM API: своё middleware, возможность self-hosting, собственный UI.
- Банк Д использует коммерческие LLM на контрактных гарантиях zero-retention, оркестрация полностью своя (несколько оркестраторов маршрутизируют между классическими моделями, агентами с tool-calling и узкоспециализированными LLM), а поверх продакшн-трафика работают собственные маленькие модели с дешёвыми асинхронными проверками.
- Банк Е гоняет self-hosted open-source LLM строго внутри периметра (Python, Kubernetes, железо NVIDIA), сознательно избегая тяжёлых фреймворков — команда считает их избыточными многослойными обёртками.

- B2B-кредитная организация строит внутренних агентов на облачной агентной платформе с собственными guardrails (не взяла даже бесплатные guardrails облачного провайдера).
- На менее зрелом полюсе: банк З собирает агентские процессы на low-code инструментах; банк Ж оценивает коммерческие наборы распознавания речи и закупает GPU под онпрем-ИИ; целевая архитектура — примерно 70% детерминированного NLU плюс небольшая LLM-составляющая: сознательный дизайн сдерживания галлюцинаций.
- Банк Б, напротив, сторонник «покупать, а не строить»: чат-боты и ассистенты поставляют вендоры, развёртывание гибридное — облако для масштабируемости, контролируемая среда для чувствительных данных.

«Строить или покупать»: дефолт определяется сегментом

Три режима, а не одна рыночная норма:

- **Строить сами** — цифровые и инженерно-ориентированные организации. Банк Д строит весь стек сам и в этой области не покупает ничего; банк Е относит почти всю техническую поверхность к самостоятельно строимой (единственная признанная покупка — кодифицированная проверка ответов по правовому/комплаенс-контенту); B2B-кредитор строит даже guardrails. Для любого внешнего вендора контроля и надзора основной конкурент в этом сегменте — внутренняя разработка.
- **Покупать модели, строить обвязку** — крупные incumbent'ы зрелых рынков. Логика банка А — время до соответствия требованиям: готовая компетенция укорачивает путь на годы по сравнению с внутренней разработкой. Банк Б отдаёт вендорам и сам агентный слой.
- **Вендорско-тендерный режим** — развивающиеся рынки. Банк Ж закупает всё через формальный тендер (3-5 месяцев плюс согласования на уровне группы); технологическая дочка банка видит ценность внешних продуктов именно в заранее встроенном локальном регуляторном контенте — регуляторика проанализирована заранее и зашита в продукт.

Вендор диалогового ИИ добавляет оговорку со стороны предложения: кодящие агенты делают заказную разработку настолько дешёвой, что организации всё чаще предпочитают строить всё сами, а функции верификации и guardrails впитываются в агентные платформы как базовая встроенная функция. Вендор тут же оговорился, сославшись на собственный опыт: соседний рынок речевой аналитики он когда-то ошибочно считал нежизнеспособным — а устойчивый спрос там держится все шесть лет.

Требования, которые финансовые организации предъявляют к технологиям и поставщикам

Требования ужесточаются по двум осям — регуляторный режим и сегмент:

- **Периметр и резидентность данных.** Самый жёсткий гейт на рынках с локализацией: банк Е называет работу строго внутри периметра главным, не подлежащим обсуждению ограничением; на рынке банка Ж облако в банковской сфере не используется вообще, а трансграничная передача банковских данных запрещена — онпрем (в контейнерах) единственная развёртываемая форма. Зрелые режимы защиты данных требуют резидентности внутри региона даже на гиперскейлерах. Банк Д обходит вопрос резидентности инженерно: контрактный zero-retention, веерное удаление по требованиям защиты данных, строгие IAM/PII-политики.
- **Сертификации и due diligence вендора.** Сертификации по кибербезопасности (PCI DSS, SOC 2) как базовый уровень у платёжного провайдера; SOC 2, сертификаты по управлению ИИ, соглашения об обработке данных, документация об ответственности и due diligence названы консультантами как блокирующие закупку — регуляторное внимание распространяется на самого вендора, а не только на продукт. Сторонний риск облачных вендоров, подключаемых к банковским ядрам, охарактеризован как дополнительная поверхность риска.
- **Гейты управления.** Многофункциональные комитеты одобрения — норма у крупных организаций: юристы, комплаенс, защита данных, инженерия, кибербезопасность (провайдер) или пятифункциональный комитет из ИТ, комплаенса, риска/модельного риска, безопасности/защиты данных и владельца бизнеса/продукта (банк Б). Наблюдение одного из консультантов о механике внедрения: запуск пилота — не узкое место; узкое место — вывод решения в прод, и для этого нужен внутренний чемпион, готовый проташить продукт через все внутренние процедуры.
- **Эксплуатационные требования.** Версионирование и откат моделей (банк А); многолетнее хранение предсказаний для повторного сканирования, поэтапные раскатки (внутренняя → ограниченная аудитория → полная) и независимая проверка модельного риска с намеренным барьером между владельцами и проверяющими моделями (банк Д); языковая компетентность как жёсткий гейт у банка Ж (98% диалогов на локальном языке, с диалектами и переключением кодов); ценится вендор-агностицизм — банки единодушно избегают зависимости от

одного поставщика, вплоть до явного страха внезапно потерять доступ к провайдеру моделей по геополитическим причинам.

4. Сценарии использования

Матрица собирает все сценарии, наблюдавшиеся или обсуждавшиеся в корпусе, по сегментам. Колонка вендора — то, что работает у его клиентов-финансовых организаций; колонка консультантов — то, что они видят у своих клиентских баз.

Сценарий	Крупный банк	Цифровой банк	Фин. B2B-компания	Вендор (клиенты)	Консультанты (клиенты)
1. Информационный/F AQ-бот (часы, продукты, инструкции)	● А, Б; у В: legacy-NLU ●, GenAI ● буксует ●	● Д, Е · ○ Ж	○ поддержка на этапе платежа	●	□ вопрос времени ниже \$5 млрд; ● выше
2. Транзакционное обслуживание с действиями по счёту (возвраты, переводы, страхование, споры)	● частично (А: разбор споров/фрода в залогиненной зоне)	● Д: 100+ действий · ○ Е: агентский бот в разработке	○ проверки баланса/платёжеспособности	● сложные расчётные сценарии	—
3. Подбор и рекомендации продуктов (карты, кредиты, ипотека)	● А — включая явно рискованные рекомендации	□	—	—	□ ключевая тревога по mis-selling
4. Проверки на фрод/скам	● А	—	● проверки контрагентов	—	—

Сценарий	Крупный банк	Цифровой банк	Фин. B2B-компания	Вендор (клиенты)	Консультанты (клиенты)
в сессии в адрес клиента					
5. Входящий голосовой ассистент	● Б: ограниченно	● Д, живой · ○ Ж — приоритет №1; у Е в планах	○ скован пресетам и	● один продвинутый банк-клиент полностью на голосе	—
6. Исходящие обзвоны ИИ (продажи/взыскание)	—	○ Ж, в следующем году	□ наблюдалось: голосовые роботы допродаются в соседнем нерегулируемом секторе	—	—
7. Копилот оператора в / подсказчик в реальном времени	● Б — по значимости наравне с чат-ботами	○ Ж: в тендере	—	—	—
8. Внутренние копилоты сотрудников и мультиагентная аналитика	● внутренние агенты А; группа Г: код/почта/исследования	● Д: исследовательские агенты по фроду/ML/AML	● мониторинг ковенант, аналитические отчёты для ~100 аналитиков	—	● копилот комплаенс-функции, которым пользуется сам консалтинг
9. Автономн	—	● Д: внутренн	○ сведение	—	—

Сценарий	Крупный банк	Цифровой банк	Фин. B2B-компания	Вендор (клиенты)	Консультанты (клиенты)
ые бэк-офисные агенты (без человека в контуре)		ие решения, под наблюдением регулятора	казначейства в ноль к D+1		
10. Речевая аналитика / пост-фактум QA диалогов	—	● Е, внешний модуль, планируется инсорсинг · ○ Ж: в тендере	—	● устойчивый спрос 6+ лет	—
11. Помощь в онбординге/KYC	—	—	□ наблюдалось вендоры автоматизации онбординга	—	□ консультанты называют критически важным
12. Ассистенты wealth/advise/private banking	□ подавлено (риск советов)	○ Д: консьерж, private banking, wealth и трейдинг	—	—	□ опасения по suitability и уязвимым клиентам

Как сценарии развиваются по сегментам

Матрица показывает устойчивую последовательность: организации начинают с информационного FAQ (строка 1), переходят к транзакционным действиям (строка 2) и только потом подходят к советнической территории (строки 3 и 12) — которую консервативные incumbent'ы сейчас прямо подавляют; в прод её вывел только один крупный банк (А). Голос (строки 5–6) следует за текстом с лагом в один-два года везде, кроме голосо-центричного рынка (Ж), где он возглавляет дорожную карту из соображений стоимости.

Выделяются два структурных контраста. Цифровые игроки эксплуатируют самое широкое и глубокое хозяйство (строки 1, 2, 5, 8, 9 в проде одновременно у одной организации — включая автономных внутренних агентов, за которыми уже начали наблюдать регуляторы). Консервативные incumbent'ы и B2B-игроки переворачивают порядок: внутренние копилоты и бэк-офисные агенты (строки 8–9) доходят до прода, пока клиентские строки пустуют, — внутреннее внедрение служит контролируемой средой для выращивания доверия к технологии. Платёжная B2B-компания демонстрирует граничное условие: там, где объём саппорта мал, а живое обслуживание — премиум, у клиентского ИИ нет бизнес-кейса вовсе — по словам её представителя, клиенты ждут человека и не примут виртуального ассистента.

5. Проблемы: классификация, рейтинг, приоритеты

Предлагаемая классификация

Проблемы, названные в корпусе, укладываются в семь классов:

- **А. Качество и правдивость ответов** — галлюцинации, фактически неверные ответы, устаревшая информация как актуальная, съезд языка, излишняя уверенность.
- **В. Масштабируемость надзора** — неспособность нынешних методов контроля (логирование, сэмплирование, ручная проверка) поспевать за объёмом диалогов; отсутствие контроля в реальном времени.
- **С. Управление и ответственность** — неясный внутренний владелец ИИ-риска, страх персональной ответственности, медленные многосторонние согласования, блокировка на уровне риск-аппетита совета директоров, незакрытые роли AI-governance.
- **Д. Комплаенс-содержание коммуникаций** — неавторизованные советы, mis-selling, вводящая в заблуждение информация о продукте/ставках, отсутствующие дисклеймеры, дискриминационный сбор данных, вред без нарушения буквы правил.
- **Е. Экономика и техническая зрелость** — компромисс «цена ответа против качества», задержка голоса, качество ASR на локальном языке, vendor lock-in и геополитический риск доступа, невозможность конвертировать legacy-ботов.
- **Ф. Безопасность** — prompt injection, джейлбрейки, утечки данных, экспозиция PII, злоупотребления со стороны злонамеренных пользователей.
- **Г. Регуляторное соответствие** — способность организации и её систем надзора демонстрируемо соблюдать регуляторные рамки: наложение нескольких режимов (защита данных, поведенческое

регулирование, модельный риск), скорость внедрения новых требований, документация и доказуемость перед надзором, подготовка к грядущему законодательству об ИИ и локальным требованиям (локализация данных, перелицензирование). В отличие от класса D (содержание отдельного диалога), класс G работает на уровне системы и процесса.

Рейтинг

Ранжировано по тому, как часто класс называли без подсказки, давал ли он реальные инциденты и насколько центральным его считали респонденты:

Ранг	Класс	База доказательств	Наиболее затронутые сегменты
1	В. Масштабируемость надзора	Назван ядром проблемы обоими крупными банками в проде: представитель банка Б открыто признаёт, что логирование, сэмплирование и ручные проверки второй линии не масштабируются с объёмом и что эффективного надзора в реальном времени нет; представитель банка А в письменном ответе без подсказки назвал главным разрывом отсутствие непрерывного,	Крупные банки в проде; платёжные провайдеры; любая организация по мере роста объёма

Ранг	Класс	База доказательств	Наиболее затронутые сегменты
2	А. Качество ответов	масштабируемо го, реального мониторинга вместо периодических проверок. Эхом — признание платёжного провайдера, что контроль для его живого агента ещё не разработан. Единственный оценённый в деньгах повторяющийся инцидентный паттерн корпуса: банк Е неоднократно сталкивался с неверными ответами бота и компенсирует клиентам последствия через постоянный механизм возвратов. Страх галлюцинаций диктует архитектуру у банка Ж (гибрид с 70% NLU) и тормозит конверсию GenAI у банка В. Самый зрелый	Все; экзистенциальн о для организаций до внедрения

Ранг	Класс	База доказательств	Наиболее затронутые сегменты
3	С. Управление и ответственность	цифровой банк (Д) переформулирует её в экономику: по его оценке, доведение частоты галлюцинаций до «одна на миллион» стоило бы порядка двадцати долларов за ответ — экономически бессмысленно.	
		Платёжный провайдер поставил её на первое место по собственному опыту: главной названа внутренняя дискуссия о распределении ответственности — функции уклоняются брать её из-за связанной экспозиции — плюс месяцы, потерянные на выбор LLM- вендора. Двухлетнее сдерживание клиентского ИИ у группы Г —	Крупные incumbent'ы, платёжные провайдеры

Ранг	Класс	База доказательств	Наиболее затронутые сегменты
		тот же класс на высоте совета директоров (риск-аппетит). Ответственность ложится на руководителя бизнес-линии (банк А).	
4	G. Регуляторное соответствие	Названо без подсказки обоими крупными банками в проде: представитель банка Б считает главным вызовом наложение нескольких регуляторных рамок — а не какой-то один акт; представитель банка А указывает на разрыв между ожиданием немедленного внедрения новых требований и многолетней внутренней реальностью — разрыв, напрямую мотивирующий их «покупать, а не строить».	Крупные банки регулируемых рынков; организации в юрисдикциях с грядущим законодательством об ИИ

Ранг	Класс	База доказательств	Наиболее затронутые сегменты
5	D. Комплаенс- содержание	Клиентская дорожная карта банка Б задана темпом вступающего комплексного законодательства об ИИ; банки развивающихся рынков готовятся к законам об ИИ до их принятия. Широко ожидается, пока редко переживается: вводящая в заблуждение информация наблюдается эпизодически (банк Б); консультанты зрелых рынков формулируют это как экспозицию на недобросовестные практики/честное кредитование (банк не должен неверно цитировать ставки или собирать запрещённые атрибуты); таксономия одного	Розничные банки регулируемых рынков; оспаривается скептиками

Ранг	Класс	База доказательств	Наиболее затронутые сегменты
		консультанта — mis-selling, suitability, уязвимые клиенты и вред без формального нарушения: клиента можно повести в неверном направлении, не нарушив ни одного регуляторного правила. Контрсигнал: представитель платёжной B2B-компании не смог вспомнить ни одного enforcement-кейса по содержанию ботов и считает категорию раздутой.	
6	Е. Экономика и зрелость	Кривая «цена-качество» (цифровой банк); задержка голоса блокирует встроенный голос; ASR с 98% локального языка как ядро препятствия у банка Ж; банк В 2+ года не может	Цифровые игроки (цена), развивающиеся рынки (язык/зрелость)

Ранг	Класс	База доказательств	Наиболее затронутые сегменты
7	Е. Безопасность	конвертировать NLU→GenAI; vendor lock-in и страхи геополитического отрезания. Реальна, но практиками ранжируется ниже всех: попытки инъекций ничего не дают атакующему там, где бот не держит РII (банк Е); закрывается IAM/zero-retention у цифрового банка; консультанты отмечают напряжение вокруг РII (банкиры хотят персональные данные для кросс-продаж) и теневое использование ИИ. Значимость резко растёт, как только агенты получают способность действовать.	Функции безопасности; растёт с агентскими возможностями

Спотлайт: проблема регуляторного соответствия (класс G)

Почему она выделена — четыре компонента. Регуляторное соответствие ведёт себя иначе, чем содержание отдельных диалогов (класс D): на кону не столько конкретный ответ бота, сколько способность организации демонстрируемо соблюдать несколько наложенных режимов.

- **Наложение рамок:** представитель банка Б назвал главным вызовом не какой-то один акт, а пересечение защиты данных, поведенческого регулирования и управления модельным риском — каждый режим предъявляет к одному и тому же ИИ-агенту свои требования.
- **Скорость внедрения:** в банке А новую регуляторную политику ожидается внедрить немедленно, тогда как внутреннее внедрение занимает годы — именно этот разрыв мотивирует покупать готовую компетенцию вместо строительства.
- **Доказуемость перед надзором:** экзаменаторы проверяют наличие оценки ИИ-рисков; регуляторы аудируют и вендора, и организацию (due diligence, соглашения об обработке данных); самый зрелый игрок наблюдает, что регуляторы поощряют контроль «человек в контуре», мониторинг в реальном времени и многолетнее хранение предсказаний.
- **Грядущее законодательство:** клиентская дорожная карта банка Б явно задана темпом вступающего в силу комплексного закона об ИИ; банк Ж поднимает поток AI-governance в преддверии закона, ожидаемого в горизонте примерно двух лет; в части юрисдикций выборки обязательного регулирования ИИ пока нет — там класс G проявляется через требования к данным (локализация, трансграничные ограничения), а не через ИИ-специфичные правила.

Оговорка: сегодня это преимущественно ожидаемая проблема — ни один респондент не назвал завершённого enforcement-действия за несоответствие ИИ-агента. При этом это единственный класс проблем с датированными форсирующими событиями (вступление законов об ИИ в силу, перелицензирование по платёжным директивам) — и потому самый вероятный кандидат на рост приоритета в ближайшие 12–24 месяца.

Инвентарь инцидентов

- **Повторяющиеся неверные ответы с финансовой потерей** — банк Е, прод: неверные ответы бота (например, неверно названная комиссия) вызывали компенсации клиентам через постоянный механизм возвратов; банк поглощает потери как операционную

стоимость. Единственный оценённый в деньгах инцидентный паттерн выборки.

- **Галлюцинации на стадии разработки** — платёжный провайдер: пойманы до прода, смягчены заземлением модели на базовые документы и поведенческие правила; затрат не возникло.
- **Хронический отказ legacy-бота** — группа Г (по рассказу советника): кнопочный бот на правилах сбоит регулярно; внутренний ИИ-инструмент однажды вернул неверные данные из-за ошибки джойна данных.
- **Репутационный паттерн «чуть не случилось» (соседняя отрасль)** — политически провокационный неверный ответ бота телеком-оператора, приведённый вендором как архетип страха развёртывания: компании боятся выставлять непредсказуемых агентов на большую аудиторию, потому что ответственность за один плохой ответ не ограничена.
- **Показательное отсутствие:** ни один респондент не смог назвать регуляторный штраф или enforcement-действие где-либо на рынке за содержание ответов ИИ-агента, и никто не оценил инцидент в деньгах за пределами описанного механизма возвратов. Сегодняшний спрос на контроль движим ожиданием, разблокировкой развёртывания и страхом за бренд — а не реализованными регуляторными потерями.

Дифференциация по сегментам

- Крупные банки острее всего чувствуют масштабируемость надзора и управление-ответственность; регуляторное соответствие названо без подсказки обоими крупными банками в проде (наложение рамок, скорость внедрения), а комплаенс-содержание коммуникаций — ожидаемая зона enforcement. Их реакции структурны: сдерживание клиентского ИИ, внутреннее внедрение первым, тяжёлые гейты перед продам.
- Цифровые банки в значительной мере свели качество ответов и безопасность к вопросу экономики: вопрос сдвигается от «возможен ли контроль вообще» к «стоимости контроля за диалог». Их ключевая тревога — компромисс качества и цены плюс vendor lock-in.
- Банки развивающихся рынков сталкиваются с качеством ответов и технической зрелостью как блокиерами до внедрения: страх галлюцинаций формирует архитектуру; качество языка и онпрем-ограничения формируют закупку.
- Финансовые B2B-компании либо чувствуют проблему управления и ответственности (вакуум ответственности у провайдера), либо не чувствуют ничего (нет ИИ — нет проблем) — при явно заявленной нетерпимости к репутационным скандалам как единственном универсальном страхе.

- Вендоры переживают проблемы рынка как сигналы спроса: страх развёртывания тормозит их клиентов; верификация запрашивается как функция платформы.
- Консультанты описывают комплаенс-содержание коммуникаций как грядущую волну (вопрос времени, не принципа), признавая, что их клиенты с меньшими активами ещё не развернулись в масштабе, где эти проблемы материализуются.

6. Регулирование, контроль и оценка качества

Регуляторные барьеры и установки — по типам режимов

- **Режим комплексного законодательства об ИИ.** Грядущий межсекторный закон об ИИ назван без подсказки представителем банка Б как фактор, напрямую формирующий ожидания (классификация высокого риска, прозрачность); тот же респондент сформулировал более глубокую проблему: главная сложность — не одно регулирование, а наложение нескольких рамок (защита данных, поведенческое регулирование, модельный риск). Темп клиентских дорожных карт явно привязан к новому закону. Показательно, что платёжный провайдер в схожем режиме вообще не упомянул этот закон: его форсирующая функция — перелицензирование платёжной деятельности по грядущей ревизии платёжной директивы, масштабная процедура, требующая заново подать на лицензию и точно декларировать деятельность, — наряду с регулированием защиты данных и доступности. Даже внутри одного режима то, какое регулирование заставляет действовать, зависит от сектора организации.
- **Секторальный поведенческий режим.** Описан консультантами как секторальная модель: в отличие от широкого закона об ИИ, здесь регулятор финансового поведения работает поверх регулятора данных — требуя секторальной кредитibility от любого вендора. Банк Д считает комплексное законодательство об ИИ излишне широким и удобным для толкования — а потому слабой форсирующей функцией — и наблюдает, что регуляторы ещё не определились, что именно регулировать, и учатся, наблюдая за организациями; отношения с регуляторами он трактует как актив (проактивно вышел к одному регулятору с предложением по моделированию поведения). Инженерный взгляд банка: регуляторы поощряют «человека в контуре», мониторинг предсказаний в реальном времени и многолетнее хранение.
- **Режим надзора за модельным риском.** Управляет руководство центрального банка по модельному риску плюс корпоративная рамка модельного риска каждого банка; риски содержания

определяет законодательство о защите потребителей (недобросовестные/вводящие практики, честное кредитование/антидискриминация); датированного дедлайна нет, поэтому внедрение добровольно и задаётся темпом конкуренции. Практическая боль — реактивная скорость: новые политики ожидается внедрять немедленно, а внутреннее внедрение растягивается на годы. Экзаменаторы проверяют, оценила ли организация риски использования ИИ.

- **Режим без специальных правил об ИИ.** Конкретного регулирования или дедлайна банковские респонденты (В, З) не назвали; регуляторика предстаёт как сложность, которую нужно инкапсулировать, — воспринимаемая ценность внешнего продукта лежит в готовой локальной регуляторной функциональности. Регуляторная срочность как движущая сила покупки на этом рынке не подтверждена.
- **Режим «данные прежде всего».** Респондент прямо говорит, что аналога комплексного закона об ИИ там не существует. Действующее ограничение — персональные данные и трансграничная передача; центральный банк проявляет активный интерес, но обязательных актов не выпустил, оставляя продвинутые банки действовать впереди регулятора по собственной инициативе.
- **Режим абсолютной локализации (банк Ж).** Законодательство об ИИ существует в проекте и ожидается в горизонте примерно двух лет; облако в банковской сфере не используется, трансграничная передача банковских данных запрещена — локализация данных абсолютна. Комплаенс-стандарты материнской группы бегут впереди локального закона; заявленная позиция банка — подготовиться до закона, а не реагировать после. Поток AI-governance ожидается в горизонте года-двух.

Внутренний контроль: кому принадлежит ИИ-риск

Воспроизводится устойчивая структура управления. Комплаенс — вторая линия защиты и не владеет контролями; контроли и бюджет у владельцев бизнеса/каналов первой линии — этот тезис наиболее остро сформулирован консультантом с опытом крупной банковской группы (в его рассказе реальный центр решений — не команда комплаенса, а человек в бизнесе, у которого в руках инструменты комплаенса и контроли) и подтверждён банком А (ответственность за отказ возлагается на руководителя соответствующей бизнес-линии). Крупные организации используют многофункциональные комитеты-гейты (ИТ, комплаенс, риск/модельный риск, безопасность/защита данных, бизнес-владение). Самый зрелый цифровой банк (Д) институционализировал независимую функцию модельного риска в роли внутреннего

регулятора — с намеренным барьером между строителями и проверяющими моделями и бременем документации на владельцах моделей. На другом полюсе, где роль AI-governance не существует, контролями по умолчанию владеет инженерия (B2B-кредитор строит свой слой управления сам). Представитель платёжного провайдера формулирует целевой подход как compliance by design — и признаёт, что соответствующий внутренний контроль для живого ИИ-агента ещё не разработан и классифицирован как будущее действие.

Как диалоги оцениваются сегодня — лестница зрелости

- **Ничего ИИ-специфичного** — только человеческие KPI контакт-центра (доля потерь, CSAT): банк Ж до внедрения.
- **Ручное сэмплирование + дашборды** — живые дашборды метрик, ручные спот-проверки общей вменяемости бота, оценки удовлетворённости в конце диалога: банк Е. LLM-as-judge там воспринимается скептически — команда не видит гарантий, что модель-судья оценит ответы лучше человека.
- **Фрагментированная проверка второй линией** — логирование, сэмплирование, ручные проверки комплаенсом, признанные немасштабируемыми: банк Б. Его самодиагноз — центральная находка отчёта: величайшая сложность не в построении ИИ, а в поддержании контроля над тем, что система говорит в живом клиентском взаимодействии.
- **Гейты перед продом + периодическая валидация** — команда точности/стресс-тестирования, обеспечивающая порог 99%+ точности (precision/recall, размеченная людьми эталонная выборка, переобучение при провале), периодическая валидация модельного риска и документация — но ничего в реальном времени и инлайн: банк А.
- **Индустриализированная непрерывная оценка** — самый зрелый цифровой банк (Д): офлайн юнит-тесты на золотых наборах для каждой версии промпта/модели; онлайн LLM-судьи, оценивающие каждый продакшн-трейс по поштучным рубрикам (следование политике, тональность), откалиброванные против собственных разметчиков с согласием $\geq 95\%$ и непрерывно переразмечаемые на дрейф; дешёвая инлайн-предпроверка на галлюцинации/фактичность плюс тяжёлая асинхронная пост-проверка на собственных моделях банка; NPS как тыловой ориентир качества.

Ключевой дефицит рынка: разрыв между уровнями 3–4 и уровнем 5. Организации с наибольшими объёмами эксплуатируют наименее масштабируемый надзор — и знают это: оба крупных банка и платёжный провайдер независимо назвали непрерывный надзор в реальном времени недостающей компетенцией, а единственный банк, у

которого она есть, строил её выделенной внутренней командой ИИ-инфраструктуры годами.

Безопасность и установки на оценку

Проверка безопасности — гейт, а не драйвер: киберсертификации — обязательный базовый уровень, директора по безопасности сидят в комитетах одобрения, а паттерн оценки клиентских чат-ботов одного из советников — офлайн-симулированные разговоры, тестирование на prompt injection, затем непрерывный онлайн-мониторинг — считается хорошей практикой. Практики ранжируют риск инъекций как управляемый там, где боты не держат персональных данных, и ожидают роста его значимости по мере того, как агенты получают способность действовать. Заявленные приоритеты надзора сходятся по сегментам на превенции выше детекции: один консультант ставит первым превентивные контроли, останавливающие проблему до клиента; представитель банка Б называет высшим приоритетом само клиентское взаимодействие, потому что именно там риск материализуется; для банка Ж ключевой вопрос — детективный контроль или превентивный: увидеть галлюцинацию недостаточно, если вмешаться невозможно. Сегодня инлайн-превентивную проверку в проде реально эксплуатирует только самый зрелый цифровой банк.

7. Кто принимает решения и кто участвует во внедрении

Как принимаются решения

Корпус выявляет двухуровневую модель решений. Стратегический уровень — разрешение вообще выпустить ИИ к клиентам — решение о риск-аппетите на уровне правления: так группа Г около двух лет держала фактический запрет на клиентский диалоговый ИИ, и пересмотр сейчас форсирует конкурентное давление, а не технологическая готовность. Тактический уровень — конкретный проект внедрения — финансирует и заказывает владелец бизнеса или канала первой линии (бизнес-/продукт-владелец, руководитель клиентского сервиса). Три сегмента подтвердили это независимо: консультант с опытом крупной банковской группы определил точку решения как бизнес-владельца контролей, а не команду комплаенса; банк А возлагает ответственность за исходы на руководителя бизнес-линии; у платёжного провайдера бюджет у бизнес-/продукт-владельца, а комплаенс играет сопровождающую, поддерживающую роль.

Вторая закономерность: чем цифровее организация, тем сильнее решение смещается к технологическим ролям. У ИИ-нативных банков покупатель — руководитель дата-саенс, Chief AI Officer, руководитель модельного риска или развития ИИ-бизнеса — а не традиционный

главный комплаенс-офицер. Там, где выделенных ролей ИИ нет, владельцем контролей по умолчанию становится инженерия (B2B-кредитор). На развивающихся рынках решение формализовано тендером (3–5 месяцев) с согласованиями материнской группы; выбор вендора ведёт функциональный владелец (руководитель клиентского сервиса) при внутреннем чемпионе ИИ в операционной функции и руководителе комплаенса как отдельном утверждающем.

Сегмент	Экономическое решение (бюджет)	Функции вето/гейтов	Характер процесса
Крупный банк	Руководитель бизнес-линии/канала; стратегическое разрешение — правление (риск-аппетит)	Пятифункциональный комитет: ИТ, комплаенс, риск/модельный риск, безопасность и защита данных, бизнес-владелец; юристы	Многосторонние согласования; персональная ответственность руководителя бизнес-линии; юристы/комплаенс владеют регуляторной рамкой
Цифровой банк	ИИ/дата-лидерство (Chief AI Officer, руководитель дата-саенс, руководитель модельного риска); у менее зрелых — руководитель клиентского сервиса	Независимая функция модельного риска в роли внутреннего регулятора (с барьером между строителями и проверяющими) ; безопасность; согласования материнской группы	Поэтапные раскаты (внутренняя → ограниченная → полная аудитория); на развивающихся рынках формальный тендер 3–5 месяцев
Фин. B2B-компания	Бизнес-/продукт-владелец; где нет выделенных ролей ИИ — инженерия по умолчанию	Комитет: юристы, комплаенс, защита данных, инженерия, кибербезопасность	Связывающее ограничение — распределение ответственности между функциями; сроки задаются согласованиями , а не разработкой
Вендор	Бизнес-спонсор	Сертифицирова	Цикл

Сегмент	Экономическое решение (бюджет)	Функции вето/гейтов	Характер процесса
(клиенты)	на стороне банка	нное (финансового класса) облако; поэтапные многострановые пилоты	развёртывания у консервативного крупного банка — до шести лет
Консультанты (клиенты)	Советническое влияние на закупку (оценка и рекомендация ПО банкам)	Аудиторы/ экзаменаторы проверяют наличие оценки ИИ-рисков; due diligence вендора (соглашения об обработке данных, сертификаты по управлению ИИ и безопасности)	Вывод клиентского ИИ требует подписи клиентского опыта, риска, комплаенса и кибербезопасности

Типичный состав проекта внедрения

- **Бизнес-/продукт-владелец** — заказывает сценарий и клиентские пути; держит бюджет.
- **Инженерия / ИИ-команда** — оркестрация, guardrails, интеграции; главный исполнитель в организациях «строим сами», с выделенной командой ИИ-инфраструктуры у самого зрелого игрока.
- **Клиентский сервис / операции** — владеет контакт-центром и KPI качества, тональностью, эскалациями; в двух банках выборки эта функция ведёт проект.
- **Комплаенс (вторая линия)** — специфицирует требования к контролю (подход compliance by design); в одной организации функция комплаенса одновременно является операционным владельцем ИИ-агента — характерное совмещение ролей ранней зрелости.
- **Модельный риск / валидация** — независимая приёмка моделей и требования к документации.
- **Безопасность и защита данных** — периметр, PII, сертификации; функция гейта, а не инициатор.
- **Юристы** — контракты с провайдерами моделей (включая zero-retention), соглашения об обработке данных.

- **Команды данных и разметки** — золотые наборы и человеческая разметка для калибровки систем оценки качества.
- **Закупка / управление вендорами** — тендер, due diligence, согласования на уровне группы.
- **Внешние участники** — платформенный вендор; профильный консультант (перевод регуляторики в технические требования, готовность вендора); материнская группа; регулятор — как наблюдатель, а у самого зрелого игрока как сторона проактивного диалога.

Следствия для проектов внедрения:

- Многосторонние согласования делают организационный риск главным риском проекта: у платёжного провайдера внутренняя дискуссия о распределении ответственности и выбор LLM-вендора съели больше ресурсов, чем сама разработка.
- Проекту нужен внутренний чемпион — как заметил один консультант, узкое место не запуск, а вывод решения в прод: без сотрудника, готового протащить продукт через внутренние процедуры, внедрение до прода не доходит, а страх менеджеров перед персональной ответственностью — реальный блокер.
- Точка входа в организацию зависит от сегмента: ИИ/дата-лидерство у цифровых игроков, бизнес-линия с одобрением комитета у крупных банков, функциональный владелец плюс тендерная процедура на развивающихся рынках.

8. Планы развития и горизонты

По сегментам

- **Крупные банки.** Банк Б планирует продвинутую помощь в приложении, кросс-канальные сценарии и больше голоса — с темпом, явно обусловленным вступающим комплексным законодательством об ИИ. Банк А не раскрыл дорожную карту, но обнаружил траекторию через логику закупки: покупка готовой компетенции, чтобы сжать многолетний цикл внутренней разработки. Банк В продолжает пока безуспешную конверсию NLU→GenAI. Группа Г, описанная одним из консультантов, выталкивается из сдерживания конкурентным давлением: в их оценке, отказ от новых инструментов больше невозможен, потому что финтехи их уже используют.
- **Цифровые банки.** Банк Д исполняет AI-first дорожную карту: ассистент как главная точка входа, динамически генерируемые интерфейсы в чате, консьерж, ассистенты private banking и wealth/трейдинга, голос уже живой, взаимодействие с регуляторами на новых рынках — катящаяся программа, а не датированный план. Банк Е выведет кнопочного бота, переведёт

практически весь консультационный поток на GenAI-бота, добавит голосовой канал, выпустит агентские действия (в разработке сейчас) и зайнсорсит речевую аналитику. У банка Ж самый конкретный календарь в корпусе: выбор платформенного вендора и онбординг управления персоналом в ближайший год, запуск голосового бота и чат-бота в следующие один-два квартала, исходящие обзвоны годом позже и комплаенс-поток AI-governance в горизонте 12-28 месяцев.

- **Прочие финансовые B2B-компании.** Платёжный провайдер будет масштабировать автоматизацию саппорта далеко за сегодняшние 5-10%, которые сам оценивает как стартовую точку, — движимый клиентским спросом на скорость — и планирует автономный агентский ИИ для сведения казначейства в конце дня; формализация внутреннего контроля для ИИ-агента признана необходимым будущим шагом, пока не профинансированным. Кредитная организация продолжает расширять внутренних агентов и строит свой слой AI-governance; клиентский ИИ не на её горизонте. Платёжная компания не планирует ИИ вообще — её премиум в живом обслуживании.
- **Вендоры.** Вендор диалогового ИИ прогнозирует массовое принятие голоса через 1-1,5 года, рекомендуя организациям начинать с чата; спрос на guardrails/верификацию видит как следующий шаг зрелости рынка. Технологическая дочка банка прорабатывает канальную дистрибуцию инструментов надзора по своей клиентской базе финансовых организаций.
- **Консультанты.** Советники описывают внедрение у клиентов как вопрос времени, а не принципа; ожидают, что инструменты надзора ускорят внедрение, снижая воспринимаемый риск, и поднимают внутренние рабочие группы по ИИ; один консультант призывает к секторальным одноюрисдикционным пилотам и регуляторному взаимодействию до запуска.

Общая временная шкала

Три перекрывающиеся волны:

- **Сейчас → 12 месяцев:** масштабирование автоматизации текстового чата и копилотов операторов; первые агентские действия по счетам у цифровых игроков; закупочные циклы (тендеры, выбор вендоров) у поздних adopters; рабочие группы по ИИ и структуры управления формируются повсюду.
- **12 → 24 месяца:** голос уходит в мейнстрим (прогноз вендора и явные планы трёх организаций сходятся на этом окне); агентские действия распространяются от цифровых игроков к incumbent'ам; приземляются запуски развивающихся рынков; комплаенс-потoki

AI-governance стартуют на развивающихся рынках по мере принятия локальных законов.

- **24 месяца и дальше:** AI-first модели банкинга (ассистент как первичный канал), автономные бэк-офисные агенты как стандартная практика и вопросы agent-to-agent/агентной коммерции (по рассказу одного респондента — самая громкая тема крупной отраслевой конференции этого года) — для которых механизмы контроля и ответственности по общему признанию ещё не определены.

Поперёк всех трёх волн в рассказах самих респондентов повторяется одна асимметрия: планы развёртывания профинансированы и датированы; планы контроля признаны и отложены — один банк открыто признаёт, что ещё не достиг стадии контроля того, как работают его агенты, понимая при этом, что необходимость придёт быстро. Организации, ушедшие дальше всех, разрешили эту асимметрию, построив надзор внутри годами; остальной рынок ещё не решил, как он это делает.