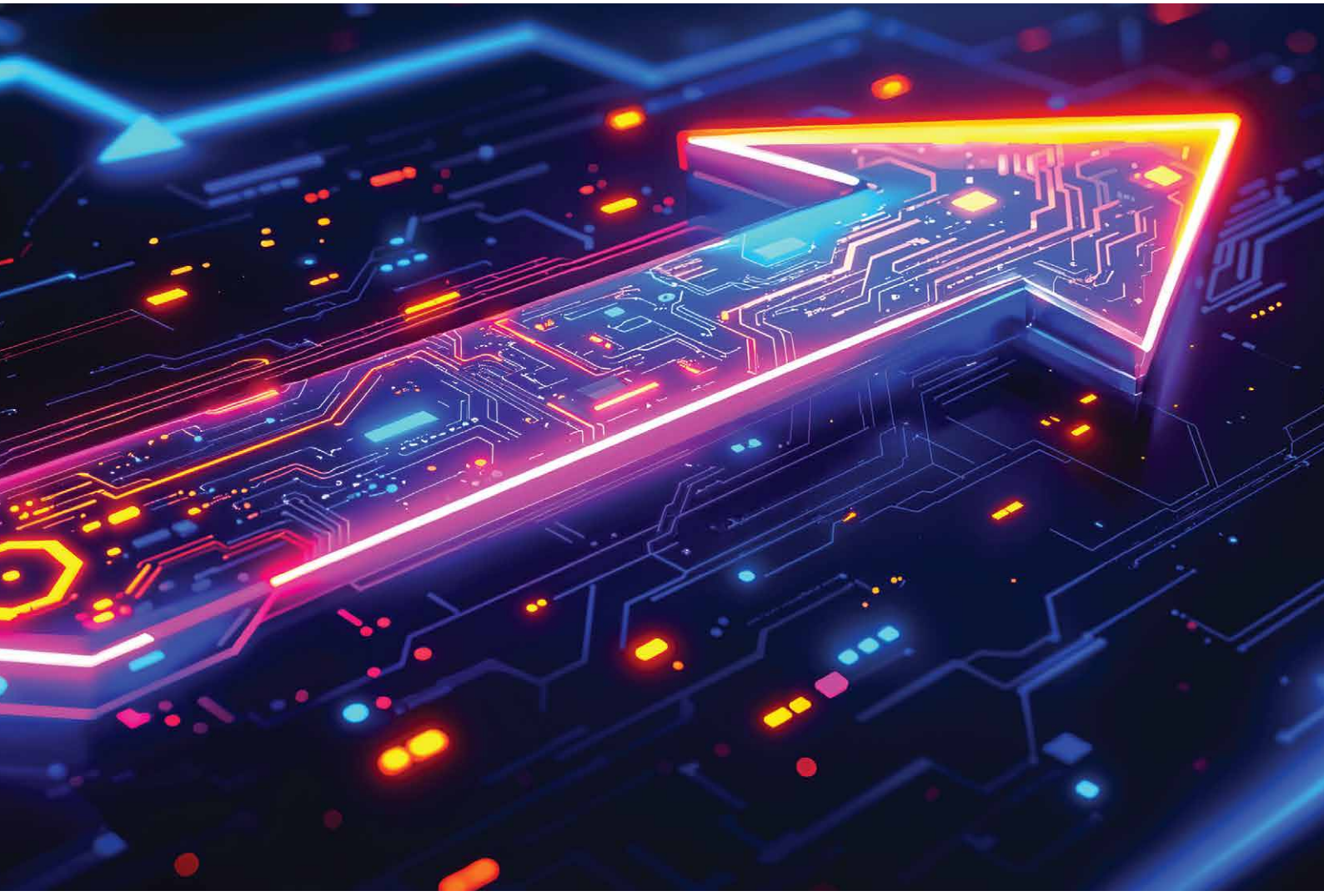
axenix

СТРАТЕГИЧЕСКИЙ ВЗГЛЯД

МУЛЬТИАГЕНТНЫЕ СИСТЕМЫ В ФИНАНСОВОМ СЕКТОРЕ

От изолированных пилотов к промышленной автономности операций



ОКТЯБРЬ 2026

Содержание

2 **Введение**

3 **Коротко о главном**

4 Экономика автономности: практика и метрики

6 Семантический фундамент: почему данных недостаточно

9 Архитектура: паттерны и промышленные требования

10 Российский ландшафт: кто задаёт темп

12 Международный опыт: уроки для российского рынка

14 Критическая оценка: риски и барьеры доверия

16 Дорожная карта: от пилота к промышленному контуру

18 Вывод для первых лиц

19 Источники

Введение

Финансовые организации переходят от изолированных чат-ботов и ассистентов к автономным операционным контурам. В основе перехода лежит мультиагентность (MAS, Multi-Agent Systems): среда, где специализированные программные агенты взаимодействуют друг с другом и самостоятельно доводят сквозные бизнес-процессы до результата.

Операции по подключению клиента (онбордингу) и сопровождению его жизненного цикла постепенно превращаются в цепочки автономных микроконтрактов между агентами. Агенты запрашивают данные, сверяют их, согласуют условия, фиксируют результат. Человек получает уведомление об исполненной операции, а его время высвобождается для работы над нестандартными случаями.

По оценке McKinsey, к 2030 году через ИИ-агентов может проходить от $3 до $5 трлн мировой потребительской коммерции в год.[10] В этой логике успех системы уже не измеряется фактом внедрения или стартовым ROI (Return on Investment, возврат на инвестиции). На первый план выходит **экономическая эффективность автономности:** какой процент операций система закрывает сама, сколько времени экономит на каждом решении, насколько снижаются операционные затраты.

Путь к таким показателям проходит через три уровня зрелости, которые индустрия описывает так:

1. Одиночный агент (Single-Agent Automation)
Один агент решает одну узкую задачу.

2. Оркестрируемая мультиагентность (Orchestrated Multi-Agent)
Оркестратор управляет специализированными агентами в рамках сквозного процесса.

3. Автономные экосистемы (Autonomous Ecosystems)
Агенты динамически распределяют задачи между собой в границах заданных бизнес-правил, состав участников процесса заранее не фиксирован.

Крупнейшие российские финансовые организации сейчас переходят от первого уровня ко второму. Это выделяет их на фоне массового сегмента и одновременно высвечивает барьеры масштабирования: незрелость архитектуры, дефицит компетенций, неготовность данных к машинному потреблению, отсутствие практик управления агентными рисками. Опрос Банка России (июль 2025 г., 252 организации) подтверждает эту картину. Процессы с высокой автономностью ИИ есть только у системно значимых банков; более 90% опрошенных организаций таких процессов не имеют.[6] Третий уровень, **по нашей оценке**, появится в российском финансовом секторе в 2028–2029 годах, сначала в комплаенс-расследованиях и договорной работе. В платёжных процессах его сдерживает рекомендация Банка России подтверждать операцию сотрудником, поэтому появление прогнозируется не ранее 2030 года.[1]

Система оценки отстаёт от практики. Пока компании ждут быстрой финансовой отдачи, возможности повышения операционной эффективности остаются вне поля зрения. Скорость принятия решений, качество продуктов, способность масштабироваться без роста штата, накопление знаний внутри организации — эти метрики показывают, какую ценность ИИ создаёт на текущем этапе зрелости.

Встроить ИИ в ключевые процессы как стандартный инструмент смогли 12% российских компаний («КОРУС Консалтинг», 2026).[5] Что отличает эти 12% от остальных и как попасть в эту группу? Ответ складывается из архитектурных решений, семантического фундамента, практик управления рисками и международных уроков.

ОЦЕНКИ АВТОРОВ[1]

Ниже расчёты и прогнозы авторов, а не данные опросов. Каждая оценка помечена в тексте и опирается на методику, изложенную в источнике 1. Все остальные цифры документа взяты из указанных источников

- **Оценка.** Оркестрируемые мультиагентные системы пилотируют не менее 2–3 из 12 системно значимых банков, входящих в перечень Банка России. К концу 2027 года их будет 7–9, а в продуктиве хотя бы на одном сквозном процессе 3–4.
- **Прогноз.** Первые автономные экосистемы (третий уровень зрелости) появятся в 2028–2029 годах в неплатёжных процессах; в платёжных — не раньше 2030 года.
- **Расчёт.** Затраты финансового сектора на агентные решения, от одиночных агентов до мультиагентных контуров, в 2026 году: 10–20 млрд руб., без учёта инвестиций Сбера в генеративные модели.
- **Оценка.** Первый процесс выходит в продуктив за 6–9 месяцев и 50–100 млн руб. Доля автономного закрытия на плато — 40–60%.
- **Оценка.** Согласованный словарь сущностей первого процесса, то есть объекты, их атрибуты и допустимые значения, занимает 30–60 строк и 3–6 недель на согласование между подразделениями.

ДИНАМИКА РЫНКА И РАЗРЫВ ВО ВНЕДРЕНИИ

- Рынок генеративного ИИ в России вырос с 13 млрд руб. в 2024 году до 58 млрд в 2025-м, примерно в 4,5 раза. На банки и страхование пришлось 12 млрд руб. (Onside, Just AI).[2]
- До полноценного внедрения дошли лишь 7–10% пилотных проектов генеративного ИИ, которые крупнейшие российские компании запустили в 2025 году; остальные остаются пилотами, меняются или закрываются («Интеллектуальная аналитика», 2026).[3]
- Технологии ИИ применяли 5,8% российских организаций по итогам 2024 года, в финансах и страховании — 7,9% (Росстат, расчёты FinExpertiza).[4]
- На финансовый сектор пришлось 62,9% всех корпоративных затрат на ИИ в стране, 56,8 из 90,3 млрд руб.[4]
- 44% компаний внедряют ИИ точечно, через отдельные эксперименты. Встроить его в ключевые процессы смогли 12%, подтвердить бизнес-эффект тоже 12% («КОРУС Консалтинг», 2026).[5]

Коротко о главном

ПОЧЕМУ СЕЙЧАС

Совпали три обстоятельства. Открытые большие языковые модели (LLM, Large Language Model) корпоративного уровня стали доступны для развёртывания в защищённом контуре. Появились первые протоколы межагентного взаимодействия, и стоимость интеграции начала снижаться, хотя единый стандарт ещё не сложился. Крупнейшие организации сектора накопили базу оцифрованных операционных данных. Пригодной для машинного потребления она становится после семантической подготовки, и это отдельная работа, которую ещё предстоит сделать.

ЧТО МЕНЯЕТСЯ

В модели «диалогового ИИ» человек формулирует запрос и получает ответ. В операционной автономности система сама доводит процесс до результата и обращается к человеку в исключительных ситуациях. Меняется не интерфейс, а организация труда в бэк-офисе и мидл-офисе.

ГДЕ МЫ НАХОДИМСЯ

Глобальные лидеры уже держат агентов в продуктиве. Nubank обрабатывает около 8,5 млн обращений в месяц, и 60% из них первыми обрабатывают языковые модели; Capital One построил собственный мультиагентный контур для клиентских и операционных задач.[11] Российские пионеры направления, Сбер и Т-Банк, внедрили мультиагентные решения в кибербезопасность, поддержку и внутреннюю разработку. Сбер за 2025 год внедрил более 900 ИИ-агентов. Заявленный банком эффект от ИИ за год (475 млрд руб.) относится ко всем видам ИИ, включая классический скоринг и ценообразование. На генеративный ИИ пришлось чуть более 50 млрд руб. при инвестициях в размере 150 млрд.[12] Остальной рынок фокусируется на локальных задачах: помощь разработчикам, автоматизация рутины, чат-боты первого поколения. Среди организаций, работающих с ИИ, агентов используют или тестируют 21% в ИТ-направлении и 16% в чат-ботах (Банк России, 2025).[6]

Разрыв между передовыми компаниями и массовым сегментом определяется не доступом к моделям — они доступны всем. Разницу создают готовность данных, зрелость архитектуры и люди, способные связать одно с другим.

СТАДИИ ВНЕДРЕНИЯ

- ИИ на постоянной основе применяет 21% организаций финансового рынка, 15% ведут пилоты, 29% планируют внедрение в течение трёх лет (Банк России, опрос 252 организаций, 2025).[6]
- Среди кредитных организаций: 24% используют, 19% пилотируют, 26% планируют внедрение в течение одного–трёх лет, 31% не планируют. Процессы с высокой автономностью ИИ есть только у системно значимых банков.[6]
- В мире ИИ-агенты в продуктиве у 23% финансовых организаций, ещё 34% пилотируют (IIF-EY, 2025).[7]
- В мире вклад ИИ в EBIT (прибыль до уплаты процентов и налогов) видят 37% компаний, значимый (свыше 5% EBIT) показывают 6%; обе доли не изменились за год (McKinsey, 2026).[8]

РИСКИ ОПЕРЕЖАЮТ ЗРЕЛОСТЬ

- Объём утечек данных через ИИ-сервисы вырос в 30 раз за 2025 год («Солар», трафик 150 компаний, 2026).[9]
- Почти 2/3 компаний не контролируют каналы утечек.[9]
- 60% не имеют формализованных политик работы с ИИ.[9]
- У 66% организаций, работающих с ИИ, нет специальных правил управления рисками ИИ (Банк России, 2025).[6]

Для перехода есть три условия:

1. Семантический фундамент

Агенты не могут договариваться без общей модели смысла. При этом строится она под конкретный процесс и в объёме словаря; централизованная программа унификации предприятия для этого не нужна.

2. Спроектированная архитектура

Выбор паттерна оркестрации задаёт прозрачность системы, её управляемость и экономику в промышленном масштабе.

3. Управление рисками как компетенция

Каскадные ошибки, стоимость вычислений и юридический статус автономных действий влияют на бизнес-модель напрямую.

2

Экономика автономности: практика и метрики

Экономический эффект мультиагентных систем возникает там, где процесс упирается в неструктурированную информацию и экспертное суждение. Розничный кредитный скоринг в России давно работает за секунды на классических моделях, и автономным агентам там добавить нечего.

Реальные кандидаты выглядят иначе:

· **Корпоративный андеррайтинг.** Анализ отчётности, договорной базы, отраслевого контекста и связанных сторон. Аналитик тратит дни на сбор и сверку, а решение должно быть объяснимо кредитному комитету.

· **Разбор клиентских диспутов.** Восстановление хронологии операций из разнородных систем, сопоставление с условиями договора и правилами платёжных систем, подготовка проекта решения.

· **Комплаенс-расследования.** Обработка сигналов AML (Anti-Money Laundering, противодействие отмыванию доходов), сбор контекста по цепочкам операций, формирование доказательной базы.

· **Сопровождение договорной работы.** Сверка редакций, выявление отклонений от типовой формы, подготовка позиции по правкам контрагента.

Эти процессы объединяет высокая доля ручного труда в повторяющихся операциях, наличие формализуемых правил и значимая стоимость ошибки.

В отчёте McKinsey по агентному ИИ в банковских операциях сквозные операции оцениваются в 60–70% базы затрат банка. Перестройка операций универсального банка от фронт-офиса до бэк-офиса силами девяти кросс-функциональных агентов даёт, по оценке McKinsey, потенциал повышения операционной эффективности на 30–40%, а централизация кредитного анализа и проверки документов снижает стоимость обработки на 20–30%.[13] На уровне банка в целом наиболее вероятный сценарий, которому McKinsey даёт 30% вероятности, даёт снижение затрат на 15–20% (Global Banking Annual Review 2025).[14] Сценарии, где агенты становятся единственным каналом обслуживания, McKinsey оценивает вероятностью ниже 5%.[14] При верном выборе процесса-кандидата затраты на пилотный контур, **по нашей оценке**, окупаются за 12–18 месяцев. Расчёт верен, если первый контур стоит 50–70 млн руб. (нижняя часть диапазона из раздела 8), а процесс до автоматизации ведёт команда около 50 человек. При стоимости ближе к 100 млн срок растёт до 24 месяцев, при команде в 30 человек до 20–28 месяцев. Капитальные вложения в инфраструктуру и расходы на эксплуатацию контура здесь не учтены.[1]

Деньги на такие проекты есть. Российский рынок ИИ по итогам 2025 года составил 316,1 млрд руб.: рост 29,7% за год и удвоение за два года, 71,9% выручки даёт корпоративный сегмент (AIANA, 2026)[15]. Бюджеты на данные и ИИ выросли у 81% финансовых организаций, у трети — более чем на 50%, и ни один банк не сообщил об отсутствии отдельного финансирования (K2Tex, 2026)[16]. Сбер планирует вложить в генеративный ИИ 350 млрд руб. в 2026 году против 150 млрд в 2025-м и чуть более 100 млрд в 2024-м — всего 600 млрд за три года. По словам Германа Грефа, речь о разработке и обучении моделей и экспериментах с ними (ноябрь 2025 г.)[12]. В мире 28% компаний тратят на ИИ более 10% ИТ-бюджета, и 60% ждут роста инвестиций в ближайший год, банки и страховщики чаще других (McKinsey, 2026)[8]. Затраты финансового сектора на агентные решения, от одиночных агентов до мультиагентных контуров, в 2026 году **мы оцениваем** в 10–20 млрд руб. Рынок генеративного ИИ в банках и страховании в 2025 году составил 12 млрд руб.[2], при росте на 20–60% в 2026 году это 14–19 млрд руб. Доля агентных решений в подобных бюджетах — 25–35%, то есть на закупки у вендоров приходится 3,6–6,7 млрд руб. Внутренняя разработка в крупных банках примерно вдвое больше внешних закупок, что и даёт итоговые 10–20 млрд руб. Инвестиции Сбера в разработку генеративных моделей сюда не входят. Оркестрируемые мультиагентные контуры занимают в этой сумме малую долю: пока их пилотируют считаные банки (см. раздел 5).[1]

КРИТЕРИИ ВЫБОРА ПРОЦЕССА-КАНДИДАТА

1. **Повторяемость и предсказуемость.** Есть чёткий чек-лист, который можно формализовать. Агент работает в заданных рамках и не изобретает подход заново.

2. **Экономическая значимость.** Высокие затраты на фонд оплаты труда и заметная цена ошибки. Перевод даже 30% таких задач в автономный режим, **по нашей оценке**, даёт измеримый эффект.[1]

3. **Готовность данных.** Данные оцифрованы, находятся в защищённом контуре и, что важнее, имеют согласованную семантику между системами-источниками.

4. **Низкий порог эскалации.** Легко задать правило передачи человеку при неуверенности агента. Процесс становится безопасным полигоном для отработки взаимодействия.

СМЕНА СИСТЕМЫ КООРДИНАТ

В спорах об окупаемости часто смешивают два разных ROI. ROI по новой выручке не работает: автономный контур в бэк-офисе продаж не добавляет, и требовать от него роста доходов в первый год бессмысленно. ROI по операционной экономии считать нужно, именно он стоит за сроком окупаемости выше. Экономия приходит позже, чем становится понятно, работает ли контур. Поэтому на ранних этапах зрелости к ROI добавляются метрики операционной автономности:

- **Доля автономного закрытия.** Процент операций, доведённых до результата без вмешательства человека.
- **Скорость принятия решений.** Время от запроса до результата.
- **Качество.** Снижение количества ошибок, возвратов и повторных обращений.
- **Масштабируемость.** Способность расти без линейного роста штата.
- **Накопление знаний.** Организация учится на каждой транзакции.

Это ведущие индикаторы. Если они растут, финансовый эффект приходит следом. Компания, которая смотрит только на прямой ROI в первые два квартала, закроет перспективный проект до того, как он окупит инфраструктуру.

Масштаб вложений растет быстрее, чем возможность оценить эффект их влияния. Количественные метрики для управления ИИ используют 24% российских финансовых организаций, около трёх четвертей обходятся без них (K2Tex, 2026).[16] Из 50 крупнейших банков мира агрегированный ROI от ИИ раскрывают 15 (Evident, 2026).[17] Согласно опросу McKinsey (1719 респондентов, 2026 г.), вклад ИИ в EBIT видят 37% организаций, значимый вклад — около 6%, и обе доли не изменились за год, хотя доля компаний, масштабирующих агентов, выросла. Каждая пятая организация уже ограничивает использование ИИ из-за операционных затрат, включая стоимость токенов.[8] Метрики процесса нужны с первого дня; ROI по экономии подтверждается позже.

3

Семантический фундамент: почему данных недостаточно

Распространённая ошибка при подготовке к внедрению MAS состоит в том, что задачу сводят к качеству данных в привычном понимании: полнота, актуальность, контроль доступа. Уровень этот необходим, но недостаточен, поскольку классическое управление данными обслуживает отчётность, где потребитель информации — человек, способный достроить недостающий контекст. У агента такой способности нет.

ПОЧЕМУ АГЕНТ НЕ ПРОЩАЕТ НЕОДНОЗНАЧНОСТЬ

Раньше эта проблема не проявлялась. В BI-контуре (Business Intelligence, бизнес-аналитика) аналитик смотрит на цифру, замечает, что она не сходится с ожидаемой, и уточняет определение метрики. Человек работал слоем разрешения неоднозначности, оставаясь незаметным. Агентная автоматизация убирает этот слой. В цепочке не остаётся ни отчёта, ни человека, который бы его прочитал. Любое расхождение в модели данных превращается в ошибку в действиях агента.

Два агента могут обмениваться синтаксически безупречными сообщениями и понимать под «клиентом», «лимитом» или «просрочкой» разные вещи, если эти понятия по-разному определены в системах-источниках. В человеческом процессе расхождение снимается на совещании. В автономном контуре оно тиражируется.

Отраслевые данные сходятся с этой логикой. Gartner прогнозирует, что к 2027 году организации, приоритизирующие семантику при подготовке данных, повысят точность агентных систем до 80% и снизят стоимость их разработки и сопровождения до 60%; в тройку опорных компонентов надёжного агентного ИИ вместе с семантикой вошли операционное состояние и происхождение данных.[18] Обратный сценарий Gartner оценивала ещё в 2025 году: свыше 40% агентных проектов будут свёрнуты или урезаны до конца 2027 года.[19] На конференции Gartner Data & Analytics 2026 семантический слой назван необсуждаемым фундаментом агентного ИИ, который нужно защищать наравне с кибербезопасностью; следующим шагом Gartner ждёт требований семантической прозрачности от регуляторов.[18]

О зрелости темы говорит и отраслевая консолидация. В сентябре 2025 года запущена инициатива Open Semantic Interchange, стандартизирующая метаданные ради переносимости семантики между аналитикой, агентами и приложениями; среди участников BlackRock версия спецификации 1.0 опубликована в январе 2026 года.[20]

Инфраструктура данных готова у большинства, сами данные — нет. 86% российских финансовых организаций оценивают свою инфраструктуру данных как полностью или частично готовую к требованиям закона, при этом 62% банков признают, что данным нужна доразметка, очистка или консолидация (K2Tex, 2026)[16]. Согласно опросу Банка России, низкое качество или нехватку данных считают барьером средней или высокой значимости 60% респондентов.[6] Мировые опросы дают ту же картину. Качество данных называют главным барьером для вывода ИИ в продуктив 79% финансовых организаций и 100% глобальных системно значимых банков и страховщиков; доступность данных отмечают 74% (IIF-EY, 2025).[7] Хранилище построено, разметка идёт, а барьер остаётся. Значит, он лежит в другом слое, в согласованном смысле сущностей.

Признак незрелости заметен без всякого аудита. Если для ответа на вопрос «сколько у нас активных клиентов?» требуется уточнять, чьё определение активности используется, организация к агентной автономности не готова. Объём накопленных данных здесь роли не играет.

ЧТО ИМЕННО СТРОИТЬ: СЛОВАРЬ ИЛИ ИНФРАСТРУКТУРУ

Тезис о важности семантики регулярно толкает компании к избыточным вложениям, поэтому стоит сразу отметить два разных архитектурных решения, которые в дискуссии обычно смешивают.

• **Канонический словарь сущностей.** Ограниченный перечень объектов, их атрибутов и допустимых значений, который передаётся модели в промпте или описании функции. «Живёт» в конфигурационном файле. Для первого процесса это, **по нашей оценке**, 30–60 строк и 3–6 недель на согласование;[1] инфраструктурных последствий нет. Срок зависит от числа систем-источников и владельцев, которым нужно договориться; объём словаря на него почти не влияет.

• **Полноценная семантическая инфраструктура.** Формальная онтология, графы знаний, механизмы логического вывода, отдельное хранилище и процесс управления. Занимает годы, требует отдельной команды и бюджета, сопоставимого с платформенным.

Простота здесь обманчива. Технически словарь — это конфигурационный файл, который пишется за пару дней. Организационно — это согласование между операционным блоком, рисками и ИТ, где у каждого своё определение просрочки, активного клиента и лимита, закреплённое в их отчётности. Именно согласование занимает недели, а в части случаев не заканчивается ничем. Мировая статистика качества данных как барьера номер один у глобальных системно значимых банков[7] описывает ровно это: технический слой давно построен, а договориться о смысле не получается.

Для первого процесса достаточно словаря. У противников формализации есть сильный исторический аргумент. Попытки построить межагентное взаимодействие на формальных онтологиях (KQML, FIPA-ACL) так и не прижились в индустрии за 20 лет, потому что создавали порог издержек, который детерминированные системы не оправдывали. Языковые модели его снимают, разбирая смысл на естественном языке. Но снимают именно порог перевода, а не согласия. Если разные подразделения дают разные определения одному и тому же событию, никакая модель этого не разрешит, потому что противоречие здесь управленческое.

Поэтому канонический словарь строится под конкретный процесс и расширяется по мере подключения следующих. Централизованная программа семантической унификации предприятия, запущенная до первого работающего контура, обычно не доходит до результата, потому что приоритеты меняются раньше, чем словарь становится полным.

КОРПОРАТИВНЫЕ ИСТОЧНИКИ И RAG-КОНТУРЫ

Технология RAG (Retrieval-Augmented Generation, генерация с опорой на извлечённые данные) позволяет агенту обращаться к проверенным источникам вместо того, чтобы полагаться только на веса модели. В корпоративном контуре это обращение к подготовленным внутренним источникам, без поиска по внешним.

Такая база знаний собирается из внутреннего документооборота: регламенты, методологии, договорная база, исторические решения. Её подготовка требует извлечения, очистки, разметки, версионирования и поддержания в актуальном состоянии. Это самостоятельный производственный процесс, у которого есть владелец и бюджет. Случайно, как побочный эффект внедрения модели, он не возникает.

АРХИТЕКТУРА ОГРАНИЧИТЕЛЕЙ

Ограничители (guardrails) — встроенные механизмы, блокирующие нежелательные действия агента на уровне промпта, промежуточного слоя и постобработки результата. Их проектируют как границы автономности, согласованные с бизнес-правилами и регуляторными требованиями до запуска системы. Ограничители, добавленные после первого инцидента, закрывают только этот инцидент.

КОНТУР ЧЕЛОВЕЧЕСКОГО КОНТРОЛЯ

Даже в зрелых мультиагентных системах человек остаётся в контуре принятия решений, а для платёжных операций его участие теперь прямо рекомендует Банк России (методические рекомендации 3-МР, июнь 2026 г.)[21].
Есть три причины:

1. **Валидация высоких рисков.** Решения с крупными финансовыми последствиями: антифрод, инвестиционный скоринг, крупные лимиты.

2. **Разрешение коллизий.** Агенты не приходят к согласованному результату или выдают противоречивые рекомендации.

3. **Эскалация нестандартных случаев.** Ситуации, требующие контекстуального суждения. Система обязана предусматривать понятный интерфейс оператора и механизм мгновенной остановки процесса.

4 Архитектура: паттерны и промышленные требования

Риски невозможно контролировать точечными исправлениями. Надёжность и безопасность предусматриваются в архитектуре изначально. Выбор паттерна оркестрации относится к стратегическим решениям и задаёт прозрачность системы, её управляемость и экономику в промышленном масштабе.

Паттерн	Применимость в финансах	Сильная сторона	Ограничение
Конвейер Pipeline	Строго последовательные процессы: сбор данных, верификация, оценка, решение	Предсказуемость. Каждый шаг фиксируется и проверяется, риск «чёрного ящика» минимален	Уязвим к каскадным сбоям. Ошибка на раннем шаге ломает цепочку, нужны компенсирующие транзакции
Центральный оркестратор Hub-and-Spoke	Многоуровневый клиентский сервис, внутренние ассистенты, сложные согласования	Высокая наблюдаемость и централизованный контроль ограничителей	Единый центр концентрирует риск. Масштабирование упирается в его пропускную способность
Совет экспертов Council of Experts	Одобрение крупных корпоративных лимитов, антифрод-мониторинг	Консенсус гасит единичные ошибки, повышает объяснимость для аудита	Растут задержка и стоимость вычислений. Для массовых операций неэффективен
Рынок задач Task Market	Динамическое распределение вычислительных ресурсов, пиковые нагрузки	Максимальная масштабируемость в неструктурированной среде	Сложен в контроле приоритетов. Без жёстких ограничителей критичные задачи вытесняются дешёвыми

Универсальной формулы не существует. Выбор определяют пять параметров: допустимая задержка, бюджет на вычисления, требуемый уровень объяснимости для аудита и регулятора, частота нештатных ситуаций, текущая зрелость команды. Устойчивее оказывается архитектура, смоделированная под конкретный профиль рисков. Копируя референсный кейс, вы наследуете и чужие значения этих пяти параметров.

О ПРОТОКОЛАХ МЕЖАГЕНТНОГО ВЗАИМОДЕЙСТВИЯ

Обмен сообщениями между агентами опирается на протоколы, из которых в публичном поле чаще других упоминаются MCP (Model Context Protocol) и A2A (Agent-to-Agent). Здесь важно удерживать правильный уровень. Протокол относится к транспортному и интеграционному слою; содержанием трансформации он не является. Стандарт в отрасли пока не устоялся, конкурирующих спецификаций несколько, а большинство промышленных внедрений сегодня работает на проприетарных механизмах оркестрации внутри выбранной платформы.

Поэтому архитектору стоит закладывать абстракцию транспортного слоя и не привязывать бизнес-логику к конкретной спецификации, которая может не пережить ближайшие два года.

СЕМЬ ТРЕБОВАНИЙ К ПРОМЫШЛЕННОЙ РЕАЛИЗАЦИИ

1. **Надёжность.** Восстановление при сбое агента без потери контекста.
2. **Согласованность состояния.** Распределённое состояние без рассинхронизации между узлами.
3. **Масштабируемость.** Горизонтальное добавление агентов под пиковую нагрузку.
4. **Наблюдаемость.** Сквозная трассировка каждого вызова: идентификатор цепочки, задержка, стоимость.
5. **Безопасность.** Изоляция агентов, ролевая модель доступа (RBAC), шифрование контекста, защита от промпт-инъекций.
6. **Управление затратами.** Контроль стоимости вычислений и автоматическое отключение при превышении бюджета.
7. **Производительность.** Заданный уровень обслуживания (SLA) по задержке на уровне оркестратора.

5

Российский ландшафт: кто задаёт темп

Разрыв между первым эшелоном и остальным рынком нарастает. Лидеры считают эффект от ИИ сотнями миллиардов рублей, тогда как большинство организаций остаётся на стадии точечных экспериментов. Опрос Банка России (июль 2025 г., 252 организации, банки в выборке покрывают 90% активов сектора) даёт первую полную картину стадий. Среди кредитных организаций ИИ на постоянной основе применяют 24%, ещё 19% тестируют пилоты, 26% планируют внедрение в течение одного–трёх лет, 31% не планируют. Страховщики впереди, 60% из них используют ИИ постоянно.[6]

Внутри группы «Используют» интенсивность разная. У системно значимых кредитных организаций (СЗКО, в перечне Банка России их 12) традиционный ИИ стоит в скоринге и риск-менеджменте в 11 из 12 организаций, а самостоятельность ИИ в скоринге физлиц, малого и среднего бизнеса приближается к 100%. Три из них применяют ИИ-агентов, в консультировании и документообороте, еще одна — для разработки и валидации других моделей. В остальных процессах агентов почти нет. Их используют или тестируют в документообороте 9% организаций, работающих с ИИ; в работе с претензиями — 5%, в аналитике и кадровых процессах — по 4%.[6] Если считать только оркестрируемые системы, где программный оркестратор ведёт минимум двух специализированных агентов по одному сквозному процессу на реальных данных, то, **по нашей оценке**, их пилотируют не менее 2–3 из этих 12 банков. К концу 2027 года мы ожидаем 7–9 таких банков, из них 3–4 с продуктивом хотя бы на одном процессе. Ассистенты разработчиков, чат-боты и роботизация одиночных операций в этот расчет не входят.[1]

Согласно опросу финансовых директоров крупнейших банков, ИИ применяют 88% (Б1 и НКР, 32 банка, 80% активов, 2026)[22]. Расхождение объясняется выборкой и определением. В выборке Банка России есть малые банки, а «использование ИИ» у Б1 включает скоринг и антифрод, которые работают больше 10 лет. Для темы этого документа важнее вторая цифра того же опроса. Формализованная стратегия развития ИИ есть у 28% банков.[22]

ЛИДЕРЫ СЧИТАЮТ В ДРУГИХ ЕДИНИЦАХ, И ЗДЕСЬ НУЖНА ТОЧНОСТЬ В ОПРЕДЕЛЕНИЯХ

Сбер за 2025 год реализовал более 700 ИИ-инициатив[23] и оценил совокупный эффект от ИИ в 475 млрд руб.; за 2020–2024 годы накопленный эффект банк оценивал в 1,3 трлн руб.[12] Это эффект от всех видов ИИ, включая классические модели скоринга, ценообразования и персонализации, которых у банка более 2000. Считается он по собственной методике банка, включает и рост доходов, и экономию.

На генеративный ИИ пришлось чуть больше десятой части, свыше 50 млрд руб., при инвестициях в него 150 млрд руб. за тот же год. Герман Греф в ноябре 2025 года прямо сказал, что вложения в генеративный ИИ пока не приносят ожидаемой прибыли.[12]
В 2026 году банк ждёт более 500 млрд руб. эффекта, в том числе около 100 млрд от генеративного ИИ.[12] В первом квартале 2026 года ИИ участвовал в решении 65% обращений клиентов Сбера в голосовых и текстовых каналах. В голосовом канале автоматизация достигла 66% против 23% в среднем по рынку, в чатах — 71% против 67% (Frank RG)[24]. В Альфа-Банке роботизированные процессы выполняют около 3,5 млн операций в месяц, что эквивалентно работе примерно 870 сотрудников (СберУниверситет, 2026)[23]. В августе 2026 года Альфа-Банк испытал на заявках торгового эквайринга мультиагентную систему из пяти агентов с разными ролями; время до первых выводов при анализе процесса сократилось с 20 до 3 дней[23].

СБЕР

Что построено	Горизонт	Источник преимущества
Более 2000 ИИ-моделей в бизнес-процессах (2024 г.); свыше 900 ИИ-агентов, которые автономно отрабатывают 70% киберинцидентов; открытая модель GigaChat 3.0 и маркетплейс цифровых агентов (2025 г.); совокупный эффект от всех видов ИИ за 2025 год — 475 млрд руб. по оценке банка, из них более 50 млрд — от генеративного ИИ[12]	Инвестиции в генеративный ИИ — 350 млрд руб. в 2026 г.; целевой эффект от ИИ — более 500 млрд руб., в том числе около 100 млрд от генеративного (оценка банка, июнь–июль 2026 г.)[12]	Экосистема (банк, электронная коммерция, B2B) и объём кросс-доменных данных

Т БАНК

Что построено	Горизонт	Источник преимущества
Платформа машинного обучения, которая обрабатывает миллионы событий в реальном времени; ИИ-ассистент разработчика на собственных моделях; прямой эффект от ИИ в операционных процессах оценивается в несколько десятков млрд руб.[23]	Пилоты бизнес-агентов	Культура «данные прежде всего» и отсутствие унаследованной архитектуры

ВТБ

Что построено	Горизонт	Источник преимущества
ИИ экономит банку около 15 млрд руб. (оценка банка на ноябрь 2025 г.)[23]	В ближайшие два года эффект может вырасти до 50 млрд руб.[23]	Масштаб корпоративного бизнеса и транзакционных данных

money

Что построено	Горизонт	Источник преимущества
Автоматизация операционных процессов: KYC (Know Your Customer, проверка клиента), расчёты, разбор диспутов	Не раскрывается	Компактность команды и скорость итераций. Ограничение: меньший объём данных

6

Международный опыт: уроки для российского рынка

СТАДИИ И ДЕНЬГИ

Мировой финансовый сектор проходит ту же воронку, только на шаг впереди. Согласно опросу IIF и EY (61 организация из восьми регионов, включая глобальные системно значимые банки, G-SIB; 2025 г.), агентный ИИ в продуктиве у 23%, ещё 34% пилотируют, 26% — планируют.[7] Среди G-SIB пилотируют 54%. Генеративный ИИ в продуктиве уже у 84% организаций, применяющих ИИ, годом ранее было 48%, а доля пилотов упала с 41% до 10%. В этой выборке за 2 года произошёл заметный сдвиг от пилотов к продуктивной эксплуатации. 96% организаций, работающих с агентами, ведут менее 50 агентных кейсов; 61% используют агентов для автоматизации операций, внешние клиентские сценарии есть у 22%.[7]

Доля агентных решений при этом растёт быстро. Среди 50 крупнейших банков мира на них пришёлся 31% публично объявленных ИИ-кейсов первого квартала 2026 года против 15% кварталом ранее (Evident, 2026 г.)[17]. Среди компаний с выручкой более 1 млрд долларов агентов масштабируют 40% против 27% годом ранее (McKinsey, 2026 г.)[8]. Но по всем отраслям в продуктиве агентов имеют 11% организаций при 38% пилотирующих (Deloitte, 2026 г.)[25]. В финансовом секторе агентный ИИ внедряют 52% организаций, до стадии масштабирования дошли 23%; финтехи впереди традиционных банков — 57% против 45% (CCAF, 2026 г.)[26].

Расходы растут вместе с внедрением. ИТ-расходы банков и инвестиционных компаний мира в 2026 году Gartner оценивает в 857,5 млрд долларов (рост 9,5%), а к 2029 году — в 1,1 трлн.[27] Все опрошенные IIF-EY организации увеличили инвестиции в ИИ в 2025 году, половина в пределах 25%, а 36% G-SIB — более чем вдвое.[7] Технологический бюджет JPMorgan Chase составляет 19,8 млрд долларов, около 10% выручки, при более чем 65 000 технологов.[28] Lloyds к концу 2026 года выделяет 1000 сотрудников на крупнейшие агентные кейсы и уже вывел в продуктив мультиагентную систему противодействия мошенничеству.[17]

ЧЕТЫРЕ ЮРИСДИКЦИИ

Четыре юрисдикции прошли разные пути внедрения ИИ в финансовом секторе. Их опыт даёт российскому рынку уроки — как позитивные, так и предостерегающие.



США:
масштаб и собственный контур



Китай:
государство как катализатор



ОАЭ:
регулятивная песочница



ЕС:
регулирование опережает внедрение

Что сделали	Урок для РФ
JPMorgan до конца 2026 года планирует запустить агентов, которые работают час-два без участия человека. Валовые продажи его подразделения управления активами и частным капиталом за 2023–2024 годы выросли на 20%; банк связывает это в том числе с ИИ-инструментами[28]	Модели доступны всем. JPMorgan не отдаёт контур агентов внешнему поставщику и первым делом решает, как выдавать агентам учётные записи и права доступа.[28] Контроль над агентами важнее выбора модели
Программа «ИИ+» Госсовета (август 2025 года) ставит цель для всей экономики: проникновение ИИ-агентов и умных устройств нового поколения выше 70% к 2027 году и выше 90% к 2030 году[29]	Государство задаёт рынку измеримую цель. В российской модели похожим рычагом остаются целевые показатели для организаций с госучастием
В финансовых зонах Дубая и Абу-Даби (DIFC и ADGM) работают регулятивные песочницы: компании тестируют новые финансовые сервисы на реальных клиентах по ограниченной лицензии. Центральный банк ОАЭ в необязательном руководстве февраля 2026 года ожидает, что полностью автономный ИИ будет ограничен процессами низкого риска; кредитные решения без возможности вмешательства человека, по его оценке, вряд ли отвечают ожиданиям надзора[30]	У Банка России есть регулятивная песочница и механизм экспериментальных правовых режимов (ЭПР)[31]. Их расширение на автономные операции позволит проверить правовой статус таких операций до появления судебных прецедентов
Регламент ЕС об ИИ (AI Act) делит системы по уровню риска. Системы оценки кредитоспособности граждан отнесены к высокому риску: до вывода на рынок они проходят оценку соответствия. В июле 2026 года ЕС перенёс срок этих требований со 2 августа 2026 года на 2 декабря 2027 года[32]	Жёсткие требования пришлось отложить на 16 месяцев. Гибкий принципно-ориентированный режим предпочтительнее копирования жёсткой классификации

Сводная матрица

- **Контроль первичен (США).** Инвестиции в собственный контур и управление агентами должны идти раньше инвестиций в модели ИИ.
- **Измеримая цель ускоряет рынок (Китай).** Целевые показатели для организаций с госучастием остаются самым прямым рычагом в российских условиях.
- **«Песочница» снижает правовой риск (ОАЭ).** Без расширения периметра регулятивной песочницы и ЭПР бизнес будет откладывать внедрение из-за юридической неопределённости.
- **Избыточное регулирование убивает скорость (ЕС).** Требуется чёткое разграничение: что подлежит предварительной оценке соответствия, а что последующему мониторингу.

7

Критическая оценка: риски и барьеры доверия

Каждый из перечисленных рисков имеет конкретную бизнес-стоимость: штрафы регулятора, операционные потери, задержку вывода продукта, репутационный ущерб. Компании, не управляющие рисками проактивно, обнаружат их в отчёте о прибылях и убытках раньше, чем успеют выстроить защитный контур.

ОПЕРАЦИОННЫЕ И ТЕХНОЛОГИЧЕСКИЕ РИСКИ

• **Каскадные ошибки.** Ошибка одного агента распространяется по цепочке и умножается. Нужны точки принудительной верификации на каждом узле.

• **Экономика вычислений.** Без контроля стоимости запросов автономный контур обходится дороже штата аналитиков. Расчёт стоимости обработки входит в бизнес-кейс с самого начала. Токены при этом дают лишь 20–25% переменных затрат агента в банковском клиентском сервисе, 70–75% приходится на человеческий контроль (McKinsey, 2026 г.)[33].

• **Управление версиями агентов.** Обновление базовой модели способно непредсказуемо изменить поведение продуктивного агента. Необходим конвейер автоматизированной доставки с регрессионным тестированием каждой версии. Практику освоили единицы российских компаний. Банки, уже работающие с агентами, главным риском называют дрейф модели — 69%, затем необъяснимость действия — 55% и ответственность за него — 47% (IIF-EY, 2025 г.)[7].

• **Кибербезопасность.** Расширение агентных полномочий увеличивает поверхность атаки. Объём утечек через ИИ-сервисы вырос в 30 раз за год.[9]

ДАННЫЕ И ИНФРАСТРУКТУРА

• **Семантическая незрелость.** Расходящиеся определения базовых сущностей между системами-источниками делают агентную оркестрацию нестабильной вне зависимости от объёма данных.

• **Инфраструктурные ограничения.** Дефицит вычислительных мощностей и ограничения на поставки оборудования удлиняют циклы развёртывания. Острая фаза, **по нашей оценке**, продлится до 2027 года.[1] В выигрышной позиции остаются игроки с собственными центрами обработки данных.

• **Зависимость от внешних моделей.** Значительная часть рынка строит решения на открытых моделях зарубежного происхождения, где актуальный лидер сегмента — семейство Qwen. Среди российских организаций, работающих с ИИ, 54% выбирают зарубежные языковые модели, из них 77% — открытые; достаточность ИИ-решений на глобальном рынке положительно оценивают 62%, на отечественном — 27% (Банк России, 2025 г.)[6]. При ужесточении ограничений возникает риск недоступности обновлений. Стратегический ответ состоит в параллельном развитии отечественных моделей (GigaChat, YandexGPT) и собственного дообучения под доменные задачи.

ОРГАНИЗАЦИОННЫЕ РИСКИ

• **Юридический статус автономных действий.** Отсутствие правовой определённости по операциям, совершённым без прямого подтверждения человеком, создаёт риск оспаривания сделок. В России рамка начала оформляться. Базовый закон о поддержке развития технологий ИИ (243-ФЗ) подписан 26 июля 2026 года и действует с 1 сентября 2026 года; часть норм вступает в силу с 1 марта 2027 года.[34] Методические рекомендации Банка России 3-МР от 16 июня 2026 года впервые систематизируют риски ИИ на финансовом рынке и для критичных процессов, в том числе платёжных операций, предусматривают подтверждение операции сотрудником.[21] Для автономных контуров это значит, что в платёжных процессах долю автономного закрытия ограничит рекомендация регулятора раньше, чем зрелость технологии. Регуляторы отстают от рынка. Только 28% финансовых организаций считают, что их надзорный орган понимает агентный ИИ (IIF-EY, 2025 г.)[7], а 48% регуляторов в мире находятся на стадии изучения ИИ или не занимаются им (CCAF, 2026 г.)[26].

• **Ловушка ROI.** Попытка измерить эффект MAS через краткосрочную выручку приводит к выводу об убыточности, поскольку скрывает операционные выигрыши. Незрелость подходов к оценке экономического эффекта заставляет сворачивать инициативы до точки окупаемости инфраструктуры. Ожидания по экономии на персонале тоже опережают факт. Год назад сокращения из-за ИИ ожидали 32% компаний, по итогам года их отметили 14% (McKinsey, 2026 г.)[8]. У Сбера генеративный ИИ в 2025 году дал более 50 млрд руб. эффекта при 150 млрд руб. инвестиций, и банк продолжает вкладывать, потому что считает эффект по всем видам ИИ на горизонте нескольких лет.[12]

• **Ловушка пилота.** Организации, запустившие 5–10 успешных пилотов, **по нашим наблюдениям**, застревают на этом уровне.[1] Причины: отсутствие платформы оркестрации, разрозненные данные, неразвитые практики эксплуатации агентов. Сильный разрыв между ожиданиями руководства и готовностью бизнеса отмечают 38% российских финансовых организаций, полное соответствие — 10% (K2Tex, 2026 г.)[16].

• **Управление рисками ИИ не выстроено.** У 66% организаций, применяющих или тестирующих ИИ, нет специальных правил управления рисками ИИ, 42% не классифицируют модели по уровню риска, 41% не ведут учёт ИИ-инцидентов; отдельная стратегия по ИИ есть у 31% (Банк России, 2025 г.)[6]. В мире 58% организаций требуют для сторонних моделей той же валидации, что и для собственных, но выполняют её 32% (IIF-EY, 2025 г.)[7].

• **Кадровый дефицит.** Дефицит ИИ-специалистов считают барьером средней или высокой значимости 74% финансовых организаций (Банк России, 2025 г.).[6] По всей экономике нехватку таких специалистов признают 79% компаний (OrionSoft, 2026 г.)[35]. Гибридных компетенций, где сходятся ИИ, разработка и информационная безопасность, на открытом рынке почти нет.

• **Доверие клиентов.** Положительно к ИИ в банковских процессах относятся 43% клиентов; доверить ИИ перевод готовы 18%, открытие вклада — 16%, оформление кредита — 12% (НАФИ, 2025 г.).[6]

8 Дорожная карта: от пилота к промышленному контуру

Прежде чем планировать сроки, топ-менеджмент отвечает на три вопроса. Какой процесс станет «первым кандидатом» при понятном эффекте и низком регуляторном риске? Кто внутри организации владеет агентной платформой: ИТ, бизнес или выделенная команда? Согласована ли у нас семантика сущностей, которые участвуют в этом процессе? Без ответов дорожная карта останется слайдом.

Работы идут четырьмя параллельными потоками. Типовые дорожные карты смешивают потоки в один список — и это ошибка, потому что у потоков разные владельцы, длительность и логика финансирования.

Горизонт планирования сознательно ограничен 3 годами. Протокольный слой межагентного взаимодействия ещё не устоялся, и планирование на 3+ года в этих условиях даёт формулировки, которые невозможно ни принять, ни отклонить.

Поток	Горизонт 1 (0–12 мес.)	Что считаем результатом	Горизонт 2 (1–3 года)	Владелец
Данные и семантика	Канонический словарь сущностей первого процесса; выписанные расхождения между системами-источниками	Определения согласованы бизнесом и ИТ письменно; список расхождений закрыт или принят как осознанный риск	Расширение словаря на смежные процессы; подготовленные базы знаний	CDO (директор по данным) / владелец данных
Архитектура и платформа	Абстракция транспортного слоя; выбранный паттерн под первый процесс; базовые ограничители	Смена спецификации протокола не затрагивает бизнес-логику; ограничители покрывают согласованный перечень запретов	Платформа эксплуатации агентов; конвейер версионирования с регрессионным тестированием	Главный архитектор
Бизнес-процессы	Один процесс в продуктивной эксплуатации	Доля автономного закрытия вышла на плато (**по нашей оценке**, 40–60%); расхождение с человеческим решением ниже согласованного порога три недели подряд	2–3 приоритетных процесса; доля автономного закрытия как управленческая метрика	Владелец процесса
Регуляторика	Внутренняя политика автономности агентов, согласованная с рекомендациями 3-МР[21]; заявка в регулятивную песочницу	Политика утверждена; юридическая позиция по статусу автономных операций зафиксирована письменно	Согласованные с регулятором сценарии; отработанные процедуры аудита	CRO (директор по рискам) / комплаенс

КАК ЭТО ВЫГЛЯДИТ НА ПРИМЕРЕ ОДНОГО ПРОЦЕССА

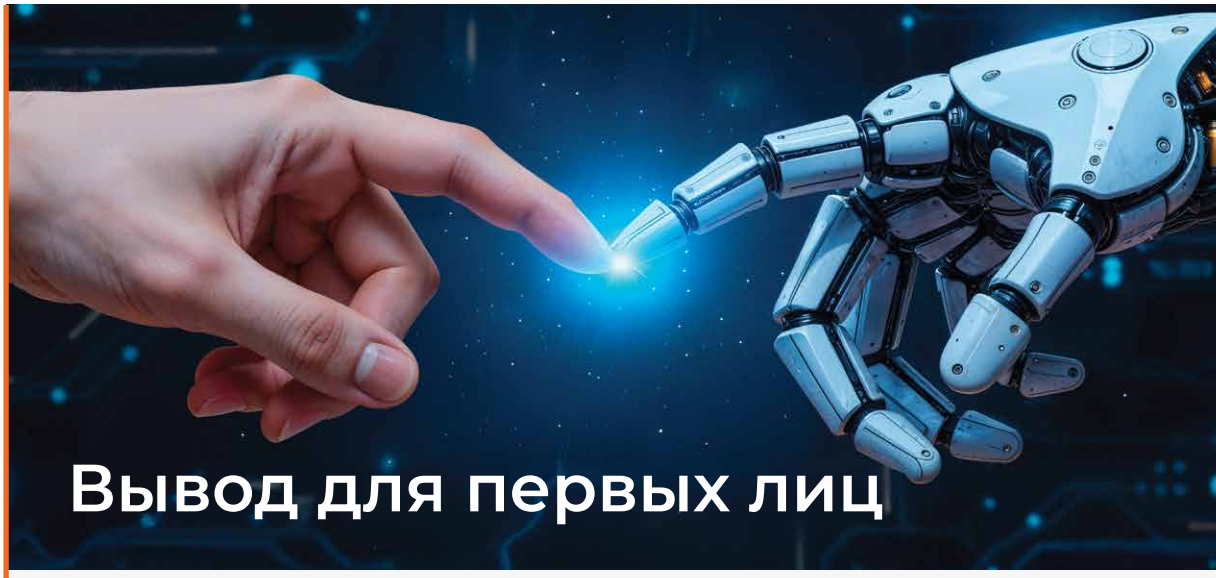
Абстрактная дорожная карта проверяется только «протяжкой» одного процесса через все потоки. Сроки этапов в таблице взяты **из нашей методики**. В таблице календарные окна, внутри которых идут работы. Этап словаря занимает 3–6 недель, но стартует не в первый день и укладывается в окно 1–3 месяца.[1] Пример построен на разборе клиентских диспутов, где высока доля ручного труда, правила формализуемы, а порог эскалации низкий.

Такой выбор процесса согласуется и с опросами. Российские организации, планирующие внедрять ИИ в ближайшие 3 года, чаще всего отдают агентам взаимодействие с клиентами, ИТ-направление, документооборот и работу с претензиями (27%, 27%, 22% и 16% соответственно; Банк России, 2025 г.).[6] Разбор диспутов лежит ровно в этой зоне.

Этап	Календарное окно	Содержание	Признак завершения
Выбор и обоснование	0–1 мес.	Проверка процесса по четырём критериям; замер объёма в штуках за месяц; замер текущего времени и стоимости обработки одного обращения	Владелец процесса назначен; исходные показатели зафиксированы в цифрах и приняты финансовой функцией
Словарь сущностей	1–3 мес.	Канонические определения для сущностей самого диспута: операция, спор, основание, срок рассмотрения, решение. 30–60 строк конфигурации за 3–6 недель, не онтология предприятия	Определения подписаны операционным блоком и ИТ; расхождения между системами представлены списком
База знаний	2–4 мес.	Правила платёжных систем, внутренние регламенты, массив исторических решений с исходами	Агент отвечает на контрольный набор вопросов с опорой на базу знаний и ссылкой на источник
Теневой режим	4–6 мес.	Агент обрабатывает реальный поток параллельно с сотрудниками, его решения не применяются, расхождения разбираются еженедельно	Расхождение с человеческим решением ниже согласованного порога три недели подряд
Продуктив с эскалацией	6–9 мес.	Агент закрывает случаи, в которых уверен, остальное передаёт человеку. Стоимость обработки отслеживается ежедневно	Доля автономного закрытия вышла на плато; инцидентов, потребовавших отката решения, нет

По нашей оценке, первый контур такого масштаба обходится в 50–100 млн руб. за 6–9 месяцев. В сумму входят 7–8 ставок команды на весь срок, вычисления для пилота, лицензии платформы и операторы теневого режима; капитальные вложения в платформу не учтены. Доля автономного закрытия выходит на плато 40–60%. Остальные случаи оставлены человеку, поскольку цена ошибки в хвосте выше экономии на нём.[1]

Работа по словарю сущностей идёт параллельно с пилотом. Расширение словаря на смежные процессы длиннее по времени и ограничивает второй горизонт. Без него масштабирование с одного процесса на портфель не состоится, каким бы успешным ни был первый контур.



Вывод для первых лиц

Гонка за агентную архитектуру в финансовом секторе началась. Модели у всех одинаковые, поэтому исход решает то, что вокруг них: согласованная семантика данных, архитектура под профиль рисков и люди, способные связать одно с другим. Масштаб разрыва виден по цифрам. Лидер рынка планирует вложить в генеративный ИИ 350 млрд руб. за один год[12], а количественные метрики для управления ИИ используют 24% финансовых организаций.[16]

Компании, которые к концу 2027 года выстроят промышленный мультиагентный контур с подготовленными данными и урегулированным правовым статусом автономных операций, получат структурное преимущество. Отыграть его будет трудно, потому что такой разрыв со временем растёт. Организации, продолжающие накапливать пилоты без масштабирования, рискуют оказаться во второй лиге к концу горизонта планирования.

Время не на стороне ожидающих.

Источники

1. Оценки авторов (число системно значимых банков с оркестрируемыми системами, сроки появления автономных экосистем, затраты сектора на агентные решения в 2026 г., стоимость и сроки первого контура, размер словаря, срок окупаемости первого контура) получены пересчётом от опубликованных данных Банка России, Onside, IIF-EY, Evident и CCAF и из практики авторов; методика расчёта доступна по запросу.

2. Onside, Just AI. Исследование рынка генеративного AI 2025–2030, декабрь 2025 г. Данные по банкам и страхованию: 3DNews по материалам «Коммерсанта», 9 декабря 2025 г.

https://onside.ru/tpost/3pb6kjns31-issledovanie-rinka-generativnogo-ai-2025
https://3dnews.ru/1133625/rinok-generativnogo-ii-v-rossii-vzletel-v-pyat-raz-za-god

3. «Интеллектуальная аналитика». Опрос около 50 крупнейших компаний, декабрь 2025 г. – февраль 2026 г.; «Ведомости», 24 марта 2026 г.

https://www.vedomosti.ru/technology/articles/2026/03/24/1184974-biznes-svernul-ili-otlozhil-9-iz-10-proektov-po-vnedreniyu-generativnogo-ii

4. FinExpertiza по данным Росстата (форма 3-информ за 2024 г.): доля организаций с ИИ и затраты на ИИ по отраслям, декабрь 2025 г.

https://www.cnews.ru/news/line/2025-12-23_tehnologii_iskusstvennogo

5. «КОРУС Консалтинг». Исследование внедрения ИИ в российских компаниях, сентябрь 2026 г.

https://www.comnews.ru/content/247177/2026-09-02/2026-w36/1018/issledovanie-korus-konsalting-pochti-70-kompaniy-ne-mogut-ocenit-effekt-vnedreniya-ii

6. Банк России. Применение искусственного интеллекта на финансовом рынке: текущий статус и условия дальнейшего развития. Доклад для общественных консультаций, ноябрь 2025 г. Опрос 252 организаций, июль 2025 г. Данные НАФИ, октябрь 2025 г.

https://www.cbr.ru/Content/Document/File/185193/Consultation_Paper_20112025.pdf
https://www.cbr.ru/press/event/?id=28102

7. IIF, EY. 2025 Annual Survey Report on AI Use in Financial Services, October 2025. 61 организация из восьми регионов.

https://www.iif.com/Publications/ID/6322/2025-IIF-EY-Annual-Survey-Report-on-AI-Use-in-Financial-Services

8. McKinsey. The state of AI in 2026: On the road to ROI, August 2026. 1719 респондентов.

https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

9. «Солар». Анализ трафика 150 российских компаний: утечки через ИИ-сервисы, февраль 2026 г.

https://www.cnews.ru/news/top/2026-02-04_sotrudniki_rossijskih_kompanij

10. McKinsey. The agentic commerce opportunity: How AI agents are ushering in a new era for consumers and merchants, 17 October 2025.

https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-agentic-commerce-opportunity-how-ai-agents-are-ushering-in-a-new-era-for-consumers-and-merchants

11. Nubank. Доклад об агентных системах в клиентском сервисе, 2025 г. (8,5 млн обращений в месяц, 60% из них первыми обрабатывают языковые модели): запись на YouTube, конспект ZenML LLMOps Database. Building AI agents for 131 million customers, март 2026 г.; arXiv 2606.08867, июнь 2026 г. Capital One. VentureBeat, 2026 г.

https://www.youtube.com/watch?v=paaOevEFNlo
https://www.zenml.io/llmops-database/building-an-ai-private-banker-with-agentic-systems-for-customer-service-and-financial-operations
https://building.nu.com/building-ai-agents-for-131-million-customers/
https://arxiv.org/abs/2606.08867
https://venturebeat.com/orchestration/why-capital-one-built-its-multi-agent-ai-platform-around-open-weight-models

Источники

12. Сбер. Инвестиции в генеративный ИИ 100 / 150 / 350 млрд руб. за 2024–2026 гг. (Г. Греф, AI Journey, 19 ноября 2025 г.; его оценка отдачи генеративного ИИ: ComNews, 11 декабря 2025 г.); прогноз эффекта на 2026 г. более 500 млрд руб., в том числе 100 млрд от генеративного ИИ (А. Ведяхин, ПМЭФ, июнь 2026 г.); годовой отчёт за 2025 г.: более 900 ИИ-агентов, эффект — 475 млрд руб.; накопленный эффект 2020–2024 гг. 1,3 трлн руб. (ноябрь 2024 г.).

https://www.tadviser.ru/index.php/Статья:Искусственный_интеллект_в_Сбербанке
https://www.interfax.ru/digital/1058901
https://m.interfax.ru/1062400
https://www.comnews.ru/content/242874/2025-12-11/2025-w50/1008/effekt-germana-grefa-550-mlrd-ozhidaniya
https://aif.ru/spief2026/iskusstvennyy-intellekt-v-biznese-v-sbere-ocenili-effekt-ot-vnedreniya-ii
https://www.kommersant.ru/doc/8692057
https://www.ng.ru/news/840140.html
https://sber.pro/publication/ekonomicheskii-effekt-ot-vnedreniya-ai-reshenii-v-sbere-previsil-1-trln-rublei-za-5-let/

13. McKinsey. AI in Asia: Reimagining banking operations through agentic AI, отчёт Operations Practice, 12 декабря 2025 г., 39 с. Основан на анализе более 600 сквозных процессов и опыте работы с более чем 1000 банков, включая свыше 300 финансовых организаций Азии.

https://www.mckinsey.com/capabilities/operations/our-insights/ai-in-asia-reimagining-banking-operations-through-agentic-ai

14. McKinsey. Global Banking Annual Review 2025 (сценарии агентного ИИ в банках), октябрь 2025 г.

https://www.mckinsey.com/featured-insights/week-in-charts/bankings-agentic-ai-opportunity

https://www.mckinsey.com/industries/financial-services/our-insights/banking-matters/agentic-ai-will-shake-up-banking-shrinking-global-profit-pools

15. AIANA. Российский рынок ИИ по итогам 2025 г.: 316,1 млрд руб., «Коммерсантъ», август 2026 г.

https://www.kommersant.ru/doc/8875821

16. K2Tex. Исследование зрелости управления генеративным ИИ в финансовом секторе (80+ обследований, 150 респондентов), август 2026 г.

https://banks.cnews.ru/news/line/2026-08-27_banki_luchshe_drugih_otraslej

17. Evident Insights. AI Use Case Trends in Banking, Q1 2026. Use Case Tracker учитывает все публично объявленные ИИ-кейсы 50 банков из индекса Evident AI Index for Banks; агрегированный ROI раскрывают 15 из них. Banking Brief, июнь 2026 г. (Lloyds).

https://evidentinsights.com/insights/banking-use-case-trends-q1-2026
https://evidentinsights.com/bankingbrief/banks-to-ai-labs-its-complicated

18. Gartner. Lack of Semantics Causes Inaccurate AI Agents and Wasted Spending (Data & Analytics Summit, London), 11 May 2026; Top Data & Analytics Predictions, 17 June 2025.

https://www.gartner.com/en/newsroom/press-releases/2026-05-11-gartner-says-lack-of-semantics-causes-inaccurate-artificial-intelligence-agents-and-wasted-spending

https://www.gartner.com/en/newsroom/press-releases/2025-06-17-gartner-announces-top-data-and-analytics-predictions

19. Gartner. Over 40% of Agentic AI Projects Will Be Canceled by End of 2027, 25 June 2025.

https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

20. Open Semantic Interchange. Анонс инициативы Snowflake, Salesforce, dbt Labs, BlackRock и др., 23 сентября 2025 г.

https://www.snowflake.com/en/news/press-releases/snowflake-salesforce-dbt-labs-and-more-revolutionize-data-readiness-for-ai-with-open-semantic-interchange-initiative/

Источники

21. Банк России. Методические рекомендации 3-МР по обеспечению информационной безопасности при разработке и применении искусственного интеллекта на финансовом рынке, 16 июня 2026 г.

https://www.cbr.ru/Crosscut/LawActs/File/12204
https://www.garant.ru/products/ipo/prime/doc/414292607/
https://expert.ru/news/tsb-rekomendoval-bankam-privlekat-cheloveka-pri-ispolzovanii-ii-v-ryade-operatsiy

22. Б1, НКР. Опрос финансовых директоров 32 российских банков (80% активов), январь–февраль 2026 г. Показатели 88% и 28% приведены в отчёте НКР «CFO 2026».

https://ratings.ru/files/research/CFO_2026.pdf
https://b1.ru/analytics/b1-banking-trends-2026-survey/

23. СберУниверситет. «ИИ-финансист: как искусственный интеллект меняет работу финансовой функции», июль 2026 г. Данные Сбера, ВТБ, Т-Банка, Альфа-Банка (по материалам ТАСС и «Банковского обозрения»).

https://finance.mail.ru/article/sber-realizoval-700-ii-initsiativ-v-2025-godu-69220198/
https://bosfera.ru/press-release/uzhe-35-40-operaciy-finansovoy-funkcii-mogut-vypolnyat-ii-agenty

Данные ВТБ: «Интерфакс», 13 ноября 2025 г.

https://www.interfax.ru/business/1057737

Пилот мультиагентной системы Альфа-Банка: пресс-релиз, «Ведомости», 18 августа 2026 г.

https://www.vedomosti.ru/press_releases/2026/08/18/alfa-bank-uskoril-poisk-insaitov-v-process-mining-s-20-do-3-dnei-s-pomoschyu-ai-agentov

24. Сбер. Итоги автоматизации клиентских обращений за I квартал 2026 г.; среднерыночные показатели по данным Frank RG. ComNews, июнь 2026 г.

https://www.comnews.ru/digital-economy/content/245835/2026-06-15/2026-w25/1012/sber-avtomatiziroval-rassmotrenie-bolee-65-klientskikh-obrascheniy

25. Deloitte. Tech Trends 2026, глава Agentic AI strategy; данные приведены по Deloitte 2025 Emerging Technology Trends study.

https://www.deloitte.com/us/en/insights/topics/technology-management/tech-trends/2026/agentic-ai-strategy.html

26. Cambridge Centre for Alternative Finance. 2026 Global AI in Financial Services Report: Adoption, Impact and Risks, май 2026 г. Опрошены участники отрасли и 130 надзорных органов.

https://www.jbs.cam.ac.uk/faculty-research/centres/alternative-finance/publications/2026-global-ai-in-financial-services-report/

27. Gartner. Forecast: Enterprise IT Spending for the Banking and Investment Services Market, Worldwide, 2023–2029, 4Q25 Update, February 2026.

https://www.gartner.com/en/documents/7431362

28. JPMorgan Chase. Технологический бюджет и агенты: Fortune, апрель 2026 г.; CNBC и InvestmentNews, июнь 2026 г.; рост валовых продаж за 2023–2024 гг.: Reuters, май 2025 г.

https://fortune.com/2026/04/29/capcom-virgin-voyages-bet-on-ai-to-reshape-gaming-and-cruise-travel/
https://www.cnbc.com/2026/06/09/jpmorgan-chase-ai-agents.html
https://investmentnews.com/transformation/jpmorgans-new-ai-agents-can-work-for-hours-without-human-input/266944
https://www.aol.com/news/jpmorgan-says-ai-helped-boost-170825172.html

29. Госсовет КНР. Руководящие указания по программе «ИИ+» (AI Plus), 27 августа 2025 г.

https://english.www.gov.cn/policies/latestreleases/202508/27/content_WS68ae7976c6d0868f4e8f51a0.html

Источники

30. DFSA. Innovation Testing Licence, регулятивная песочница DIFC. Kayrouz & Associates. Fintech Sandbox UAE: How DIFC And ADGM Approve Testing, апрель 2026 г. Hadef & Partners. AI in Banking & Finance in the UAE (о руководстве Центрального банка ОАЭ, февраль 2026 г.), апрель 2026 г.

https://www.dfsa.ae/FinTech

https://www.kayrouzandassociates.com/insights/regulatory-sandbox-innovation-testing-licence-difc-adgm-2026

https://hadefpartners.com/news-insights/insights/ai-in-banking-finance-in-the-uae-2026-legal-regulatory-considerations/

31. Банк России. Регулятивная «песочница» (с апреля 2018 г.; ИИ в числе приоритетных направлений).

https://www.cbr.ru/fintech/regulatory_sandbox/

32. Регламент (ЕС) 2024/1689 об искусственном интеллекте, пункт 58 преамбулы и приложение III. Регламент (ЕС) 2026/1744 (Digital Omnibus on AI), действует с 27 июля 2026 г.

https://artificialintelligenceact.eu/recital/58/

https://natlawreview.com/article/eu-digital-omnibus-ai-enters-force

33. McKinsey. Where AI agents pay off: a practical guide to the economics of agentic workflows, August 2026.

https://www.mckinsey.com/capabilities/quantumblack/our-insights/where-ai-agents-pay-off-a-practical-guide-to-the-economics-of-agentic-workflows

34. Федеральный закон от 26 июля 2026 г. № 243-ФЗ «О поддержке развития технологий искусственного интеллекта в Российской Федерации». Принят Госдумой 8 июля 2026 г., действует с 1 сентября 2026 г., отдельные положения с 1 марта 2027 г.

https://www.consultant.ru/document/cons_doc_LAW_540336/

https://www.garant.ru/news/2160376/

35. OrionSoft. Индекс готовности ИИ-инфраструктуры российских организаций, I полугодие 2026 г.

https://www.ixbt.com/news/2026/07/29/v-rossii-ne-hvataet-specialistov-po-ii-i-serverov-s-graficheskimi-processorami-dlja-raboty-jetih-tehnologii-soglasno.html

Точка зрения

Мультиагентные системы в финансовом секторе

От изолированных пилотов к промышленной автономности операций

Мы готовы поделиться выводами экспертов и помочь адаптировать материалы публикации под задачи вашего бизнеса. Вы можете связаться с командой экспертов AXENIX по всем вопросам с помощью электронной почты: **info-russia@axenix.pro.**

АВТОРЫ:

Антон Пащенко
Директор департамента «Финансовые организации»

Вероника Чуфистова
Эксперт практики «Финансы»

Все исследования и обзоры AXENIX — **на нашем сайте**

О компании
AXENIX — российская компания, предоставляющая широкий спектр профессиональных услуг в области цифровых сервисов и технологий. В офисах и центрах разработки в Москве, Твери, Ростове-на-Дону, Краснодаре и Санкт-Петербурге работают более 2 800 сотрудников. Благодаря сочетанию уникальных знаний, опыта и компетенций в различных отраслях компания может предложить услуги в области стратегии и бизнес-консалтинга, технологические решения и операции, направленные на цифровизацию бизнеса.