

Душкин Р. В.
ООО «А-Я эксперт»

Методика оценки стоимости и эффективности внедрения генеративных систем искусственного интеллекта

*Практические рекомендации для
государственных и корпоративных заказчиков*



Москва 2025

Об авторе



Душкин Роман Викторович — генеральный директор ООО «А-Я эксперт», старший преподаватель кафедры 22 «Кибернетика» НИЯУ МИФИ, руководитель образовательной программы «Искусственный интеллект».

Автор более 20 книг и 100 научных статей по технологиям искусственного интеллекта, разработчик образовательных программ для корпоративного сектора.

Имеет практический опыт создания систем ИИ для транспортной отрасли и государственных нужд, включая автоматизированные системы поддержки принятия решений и интеллектуальные транспортные системы для российских городов.

Рецензент: Забегалин Евгений Викторович, кандидат технических наук, ведущий научный сотрудник ФГБУ «4 ЦНИИ» Минобороны России.

О компании

ООО «А-Я эксперт» — российская ИТ-компания, специализирующаяся на разработке и внедрении систем искусственного интеллекта. Компания предоставляет услуги технико-экономического анализа, проектирования архитектуры и консультирования по вопросам внедрения ИИ-решений для государственного и корпоративного секторов.

Основные направления деятельности включают создание интеллектуальных систем, математическое моделирование и формирование стратегий цифровой трансформации на базе технологий искусственного интеллекта.

Аннотация

Настоящий аналитический отчёт представляет комплексную методику оценки стоимости и эффективности внедрения генеративных систем искусственного интеллекта, разработанную с учётом специфики российского рынка и требований нормативного регулирования.

Документ основан на референтной модели жизненного цикла проекта создания систем ГенИИ в соответствии с российскими стандартами и включает детальный анализ совокупной стоимости владения, различных архитектурных решений, моделей развёртывания и экономических показателей эффективности.

Отчёт предоставляет практические инструменты для планирования проектов внедрения, включая методики расчёта первоначальных затрат и операционных расходов, управления скрытыми затратами и рисками, а также критерии выбора поставщиков и технологий.

Особое внимание уделено сравнительному анализу облачных и локальных решений с расчётом точек безубыточности и долгосрочных прогнозов затрат.

Материал адресован руководителям проектов, финансовым директорам, главным инженерам и специалистам по закупкам в государственном и корпоративном секторах, принимающим решения о внедрении ГенИИ-систем и нуждающимся в обоснованных методологических подходах к оценке экономической целесообразности таких инвестиций в условиях высокой технологической неопределённости.

Отказ от ответственности

Настоящий аналитический отчёт подготовлен исключительно в информационных целях и представляет собой обобщение публично доступных данных и экспертных оценок на момент его создания.

Авторы и ООО «А-Я эксперт» не несут ответственности за точность, полноту или актуальность представленной информации, а также за любые прямые или косвенные убытки, которые могут возникнуть в результате использования данного отчёта или принятия его в качестве основы управленческих решений.

Все расчёты, оценки и рекомендации носят справочный характер и не являются индивидуальными инвестиционными советами или гарантиями достижения определённых финансовых результатов.

Читатель применяет изложенную методику на свой страх и риск и должен самостоятельно проверять всю представленную информацию, адаптировать её к специфике своей организации и при необходимости получать профессиональные консультации специалистов в области финансов, права и информационных технологий перед принятием каких-либо инвестиционных или технических решений.

Ключевые выводы исследования

Исследование показывает, что внедрение генеративных систем искусственного интеллекта требует комплексного подхода к планированию затрат и рисков.

Совокупная стоимость владения ГенИИ-системами складывается из четырёх основных компонентов: капитальных затрат на оборудование и лицензии, операционных расходов на поддержку и энергопотребление, скрытых затрат на обучение персонала и интеграцию, а также резервов на покрытие непредвиденных расходов.

Анализ различных архитектурных решений выявил, что выбор между облачными и локальными моделями развёртывания критически зависит от масштаба проекта, стабильности нагрузки и регуляторных требований, при этом точка безубыточности локальной инфраструктуры достигается через 6 — 12 месяцев непрерывной эксплуатации.

Стратегические рекомендации

Для успешного внедрения ГенИИ-систем организациям рекомендуется следовать поэтапной методологии, начиная с пилотных проектов и постепенно масштабируя решения после подтверждения их эффективности.

Критически важным является правильный выбор поставщика на основе технических, коммерческих и репутационных критериев, а также создание системы управления рисками, учитывающей специфику ИИ-проектов.

Особое внимание следует уделить управлению скрытыми затратами, которые могут составлять до 80% от общего бюджета проекта, и формированию команды с необходимыми компетенциями в области MLOps и архитектуры ИИ-систем.

Экономическое обоснование инвестиций

Экономическая эффективность ГенИИ-проектов характеризуется средним возвратом инвестиций на уровне 350 % за трёхлетний период при правильном планировании и реализации.

Для корпоративных заказчиков ключевыми показателями являются ROI, NPV и IRR, в то время как государственные организации должны сосредоточиться на достижении целевых показателей и социально-экономическом эффекте.

Срок окупаемости варьируется от 6 до 27 месяцев в зависимости от выбранной модели развёртывания и специфики проекта.

Для минимизации рисков превышения бюджета рекомендуется закладывать резерв в размере 20 — 25 % от плановых затрат и использовать систему KPI для мониторинга эффективности на всех этапах жизненного цикла системы.

Введение

Генеративный искусственный интеллект и большие языковые модели представляют собой не революционную «серебряную пулю» для решения всех бизнес-задач, а закономерный этап эволюции инструментов автоматизации. Подобно тому, как в своё время электронные таблицы автоматизировали рутинные вычисления, а системы планирования ресурсов предприятия оптимизировали управление процессами, современные ГенИИ-системы расширяют границы автоматизации в область когнитивных функций.

Эти технологии позволяют передать машинам выполнение задач, которые ранее требовали исключительно человеческого интеллекта: понимание естественного языка, генерацию текстов, анализ неструктурированных данных и принятие решений на основе неполной информации. Однако, как и любой инструмент автоматизации, ГенИИ требует трезвого экономического обоснования и системного подхода к внедрению.

Эйфория от впечатляющих демонстраций возможностей ChatGPT или Midjourney не должна заслонять фундаментальные принципы управления ИТ-проектами: тщательное планирование затрат, анализ рисков, поэтапное внедрение и измерение результатов.

В этом контексте настоящий аналитический отчёт предоставляет практические инструменты для оценки экономической целесообразности и планирования проектов внедрения ГенИИ-систем, помогая руководителям принимать обоснованные решения на основе конкретных данных, а не технологических обещаний.

Особую актуальность эта методология приобретает в условиях российского рынка, на котором организации сталкиваются с дополнительными вызовами: необходимостью соответствия требованиям информационной безопасности, ограничениями доступа к зарубежным облачным платформам и потребностью в импортозамещении критических компонентов.

Вместе с тем, российские компании имеют возможность использовать отечественные решения в области ИИ, что может обеспечить большую независимость и контроль над технологическими процессами при правильном планировании инвестиций.

1 Методология оценки совокупной стоимости владения



1.1 Референтная модель жизненного цикла проекта создания системы ГЕНИИ

Нормативная база построения модели жизненного цикла

Построение референтной модели жизненного цикла проекта создания системы генеративного искусственного интеллекта основывается на комплексе российских стандартов, обеспечивающих системный подход к управлению проектами автоматизированных систем:

- **ГОСТ Р 59793—2021** «Автоматизированные системы. Стадии создания» определяет основные стадии создания автоматизированных систем и содержание работ на каждой стадии.
- **ГОСТ Р 58302—2018** «Управление стоимостью жизненного цикла» устанавливает принципы и методы управления стоимостью на всех этапах жизненного цикла.
- **ГОСТ Р 57193—2016** «Системная и программная инженерия. Процессы жизненного цикла систем» определяет процессы жизненного цикла для систем, созданных человеком.
- **ГОСТ Р ИСО/МЭК 12207—2010** «Системная и программная инженерия. Процессы жизненного цикла программных средств» устанавливает общую основу для процессов жизненного цикла программных средств.

- **ГОСТ Р 56923—2016** «Системная и программная инженерия. Управление жизненным циклом. Часть 3. Руководство по применению ИСО/МЭК 12207» предоставляет руководство по применению процессов жизненного цикла.

Стадии жизненного цикла системы ГенИИ

На основании указанных стандартов выделяются следующие основные стадии жизненного цикла системы ГенИИ:

Стадия 1. Формирование бизнес-требований к системе ГенИИ

- Обследование объекта автоматизации
- Структурирование и документирование бизнес-требований к системе, утверждение их владельцами бизнеса

Стадия 2. Разработка концепции системы ГенИИ

- Разработка функциональной архитектуры системы
- Разработка альтернативных вариантов ИТ-архитектуры системы и их сравнительная технико-экономическая оценка
- Разработка предложений владельцам бизнеса по выбору предпочтительного архитектурного решения системы
- Представление концепции системы владельцам бизнеса

Стадия 3. Техническое задание на создание системы ГенИИ

- Выбор владельцами бизнеса предпочтительного архитектурного решения системы
- Составление технического задания на создание системы
- Утверждение владельцами бизнеса технического задания на создание системы
- Утверждение владельцами бизнеса бюджета проекта по созданию системы
- Заключение подрядного контракта на создание системы

Стадия 4. Технический проект системы ГенИИ

- Организация проектной команды и проектного офиса
- Планирование работ, сроков и ресурсов технического проекта
- Разработка технических, программных и организационных решений системы

Стадия 5. Рабочий проект системы ГенИИ

- Планирование работ, сроков и ресурсов рабочего проекта
- Производство компонентов системы
- Разработка рабочей и эксплуатационной документации

Стадия 6. Ввод системы ГенИИ в действие

- Подготовка объекта автоматизации

- Подготовка бизнес-пользователей и эксплуатационного персонала системы
- Поставка комплектующих изделий системы
- Строительно-монтажные работы
- Пусконаладочные работы
- Строительно-монтажные работы
- Пусконаладочные работы
- Предварительные испытания системы
- Опытная эксплуатация системы
- Приёмочные испытания системы

Стадия 7. Сопровождение системы ГенИИ

- Постгарантийное обслуживание
- Оперативная модернизация системы
- Перетренировка и адаптация моделей

Стадия 8. Вывод системы ГенИИ из эксплуатации

Стадии жизненного цикла системы ГенИИ

Жизненный цикл систем генеративного искусственного интеллекта имеет специфические особенности, отличающие его от традиционных автоматизированных систем:

1. Итерационность процесса разработки — необходимость многократных циклов обучения, тестирования и корректировки моделей
2. Непрерывность обучения — потребность в постоянной перетренировке моделей для поддержания актуальности

3. Высокая зависимость от качества данных — критическое влияние качества обучающих данных на эффективность системы
4. Эволюционный характер развития — необходимость постоянной адаптации к изменяющимся условиям и требованиям

Распределение затрат по стадиям жизненного цикла

Анализ затрат по стадиям жизненного цикла системы ГенИИ показывает следующее распределение:

- Формирование требований: 8 — 12 % от общих затрат
- Разработка концепции: 15 — 20 % от общих затрат
- Техническое и рабочее проектирование: 25 — 30 % от общих затрат
- Ввод в действие: 35 — 40 % от общих затрат
- Сопровождение: 15 — 25 % от общих затрат ежегодно

Представленная референтная модель служит основой для последующего расчёта совокупной стоимости владения и планирования проектов внедрения ГенИИ-систем.

1.2 Теоретические основы расчёта стоимости владения

Определение понятий и концептуальные основы

Совокупная стоимость владения (ССВ, или англ. TCO — Total Cost of Ownership) представляет собой финансовую оценку, предназначенную для определения

прямых и косвенных затрат на приобретение, владение и эксплуатацию продукта или услуги [Mazur & Wasiak, 2018]. В контексте генеративных систем искусственного интеллекта ССВ охватывает все расходы, связанные с покупкой, развёртыванием, управлением, использованием и выводом из эксплуатации ИИ-активов на протяжении всего их жизненного цикла [Stroparo & Floriani, 2024].

Согласно определению, предложенному аналитиком компании Gartner Group Биллом Кирвином, ССВ представляет собой *«общую стоимость приобретения, использования, управления и вывода актива в течение всего его жизненного цикла»* [ТСО-2023]. Это определение подчёркивает комплексный характер оценки, выходящий за рамки первоначальной цены покупки.

В российской терминологии совокупная стоимость владения определяется как *«общая величина целевых затрат, которые вынужден нести владелец с момента начала реализации вступления в состояние владения до момента выхода из состояния владения и исполнения владельцем полного объёма обязательств, связанных с владением»* [ГОСТ 58535].

Развитие концепции и исторический контекст

Концепция ССВ впервые была формализована Министерством обороны США для оценки общих затрат, связанных с оборонными программами, что привело к публикации военного стандарта в конце 1990-х годов [ТСО-2023]. В сфере информационных технологий англоязычный термин ТСО был популяризирован компанией Gartner Group в 1987 году, первоначально как средство расчёта стоимости

владения компьютером на платформе Wintel [Mieritz & Kirwin, 2005].

Благодаря компании Interpose, основанной в 1994 году, методика переросла в принципиально новую модель анализа финансовой стороны использования информационных технологий. На протяжении последующих лет различными компаниями велись работы по изучению проблемы определения затрат на информационные системы, что привело к появлению схожих методик: истинная стоимость владения (англ. RCO — *Real Cost of Ownership*), совокупная стоимость владения приложениями (англ. TCA — *Total Cost of Application Ownership*) и других.

Развитие концепции и исторический контекст

1) ISO 15686 — Стандарты планирования срока службы

Международная организация по стандартизации разработала серию стандартов ISO 15686 «Здания и построенные активы. Планирование срока службы», включающую специализированную часть 5, посвящённую оценке стоимости жизненного цикла (*Life-cycle costing*). Стандарт ISO 15686-5:2017 «Здания и построенные активы. Планирование срока службы.

Часть 5. Оценка стоимости жизненного цикла содержит рекомендации по проведению анализа стоимости жизненного цикла зданий, построенных активов и их частей [ISO 15686-5:2017].

Принципы, заложенные в серии стандартов ISO 15686, могут быть адаптированы для оценки стоимости владения ИИ-системами, поскольку они предусматривают:

- Определение периода анализа
- Учёт всех релевантных затрат на протяжении жизненного цикла
- Дисконтирование будущих затрат к текущей стоимости
- Анализ неопределённости и рисков

2) ГОСТ Р 27.202—2012 и другие российские стандарты

В России действует **ГОСТ Р 27.202—2012** *«Надёжность в технике. Управление надёжностью. Стоимость жизненного цикла»*, который содержит общую концепцию оценивания стоимости жизненного цикла продукции [ГОСТ 27.202]. Стандарт уделяет основное внимание затратам, связанным с надёжностью продукции, и включает шесть основных стадий жизненного цикла:

- Выработка концепции и установление технических требований
- Разработка
- Производство
- Установка, монтаж и наладка
- Эксплуатация, техническое обслуживание и ремонты
- Списание и утилизация

Дополнительно применяется **ГОСТ Р 58302—2018** *«Управление стоимостью жизненного цикла»*, который устанавливает номенклатуру показателей для оценивания стоимости жизненного цикла [ГОСТ 58302].

3) ГОСТ Р МЭК 60300-3-3—2021

Важным российским стандартом является ГОСТ Р МЭК 60300-3-3-2021 «Менеджмент надёжности», который определяет подходы к анализу стоимости жизненного цикла. Стандарт подчёркивает, что стоимость жизненного цикла включает только элементы затрат, которые могут быть материальными или нематериальными, но исключает доходы [ГОСТ 60300].

Адаптация международных подходов для российского рынка

Особенности применения в российских условиях. Адаптация международных методологий оценки ССВ для российского рынка требует учёта следующих специфических факторов:

- *Нормативно-правовые требования:* необходимость соответствия российскому законодательству о персональных данных и требованиям информационной безопасности.
- *Особенности государственного заказа:* специфические требования к прозрачности затрат и процедурам закупок в государственном секторе.
- *Валютные риски:* необходимость учёта колебаний валютных курсов при оценке долгосрочных проектов.

Методологические принципы адаптации. При адаптации международных стандартов для российского рынка следует руководствоваться следующими принципами:

1. Системный подход: комплексная оценка всех компонентов стоимости с учётом их взаимосвязи.

2. Временная перспектива: учёт изменения стоимости денег во времени через механизмы дисконтирования.
3. Управление неопределённостью: применение вероятностных методов для оценки рисков
4. Прозрачность методологии: возможность верификации и воспроизведения результатов оценки

Структурная схема компонентов стоимости владения

Для наглядного представления структуры ССВ ГенИИИ-системами предлагается следующая схема:

Совокупная стоимость владения ГЕНИИ-системами

CAPEX

(Капитальные затраты)

- Аппаратное обеспечение
- Программное обеспечение
- Инфраструктура
- Первоначальное развертывание

OPEX

(Операционные затраты)

- Энергопотребление
- Персонал
- Техническое обслуживание
- Поддержка и обновления

Скрытые затраты

(Непрямые затраты)

- Обучение персонала
- Интеграция с системами
- Простои при внедрении
- Адаптация процессов

Затраты на риски

(Резервы)

- Непредвиденные расходы
- Превышение бюджета
- Задержки проекта
- Технические сбои

$$\text{ССВ} = \text{CAPEX} + \text{OPEX} + \text{Скрытые затраты} + \text{Затраты на риски}$$

Схема демонстрирует иерархическую структуру компонентов ССВ, включающую три основные категории:

- Прямые затраты — затраты, которые могут быть непосредственно отнесены к ИИ-системе
- Косвенные (скрытые) затраты — затраты, которые возникают опосредованно в процессе внедрения и эксплуатации
- Затраты жизненного цикла — затраты, распределённые по временным этапам существования системы

Формула расчёта совокупной стоимости владения

Базовая формула ССВ может быть представлена в следующем виде:

$$\text{ССВ} = \text{Стоимость покупки} + (\text{Эксплуатационные расходы} \times \text{Время})$$

Для более детального анализа применяется расширенная формула:

$$\text{ССВ} = \text{CAPEX} + \text{OPEX} + \text{Скрытые затраты} + \text{Затраты на риски}$$

где:

- CAPEX — капитальные затраты на приобретение и первоначальное развёртывание

- ОРЕХ — операционные затраты на эксплуатацию и поддержку
- Скрытые затраты — не прямые затраты на обучение, интеграцию, простои
- Затраты на риски — резервы на покрытие непредвиденных расходов

Представленная формула обеспечивает комплексную оценку всех аспектов стоимости владения ГенИИ-системами и служит основой для принятия обоснованных инвестиционных решений в российских условиях.

1.3 Модель расчёта первоначальных затрат

Учёт затрат на разработку концепции системы ГенИИ

В соответствии с референтной моделью жизненного цикла (см. п. 1.1), первоначальные затраты должны включать расходы на стадии формирования требований и разработки концепции системы ГенИИ:

1) Затраты на формирование требований:

- Обследование объекта автоматизации и анализ бизнес-процессов
- Работы по формированию концепции системы
- Разработка технического задания
- Согласование и утверждение проектной документации

2) Затраты на разработку концепции:

- Разработка предварительного проекта
- Выбор архитектурного решения и технических средств
- Анализ и подготовка данных для обучения моделей
- Планирование ресурсов и составление календарного плана

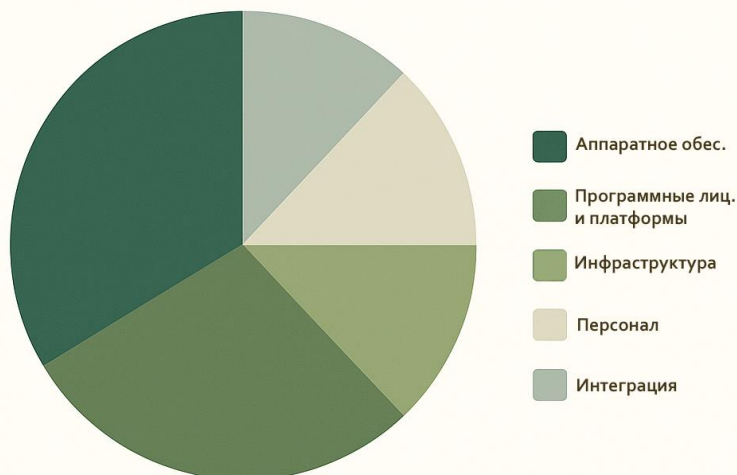
Перечисленные затраты составляют 23 — 32 % от общих первоначальных затрат проекта и критически важны для успешной реализации системы ГенИИ.

Концептуальные основы модели

Модель расчёта первоначальных затрат на генеративные системы искусственного интеллекта представляет собой структурированный подход к оценке капитальных расходов (CAPEX), необходимых для развёртывания ГенИИ-решений. Эта модель базируется на анализе четырёх основных категорий затрат, каждая из которых имеет специфические характеристики и влияет на общую стоимость проекта [TAdviser-GenAI].

Первоначальные затраты включают в себя все расходы, понесённые на этапе планирования, закупки оборудования, лицензирования программного обеспечения, создания инфраструктуры и формирования проектной команды до момента ввода системы в эксплуатацию. Согласно международной практике, первоначальные затраты составляют от 60 % до 80 % от общей совокупной стоимости владения в первые три года эксплуатации [Upperator-2023].

Структура первоначальных затрат на ГенИИ-системы



Структура первоначальных затрат на генеративные ИИ-системы

Компонентная структура первоначальных затрат

1) Аппаратное обеспечение

Аппаратное обеспечение представляет наиболее существенную статью первоначальных затрат, составляя в среднем 35 % от общего бюджета проекта [Cloud4Y-2019]. Основными компонентами данной категории являются:

- Графические процессоры (GPU) — критически важный компонент для обучения и функционирования генеративных моделей. Стоимость профессиональных GPU варьируется в широких пределах: NVIDIA Tesla P100 8GB оценивается в 17 664 рубля в месяц при аренде, NVIDIA Tesla V100 1GB — 18 341 рубль в месяц, а топовые решения NVIDIA H100 достигают 154 624 рубля в месяц. При покупке одна карта H100 может стоить до 25 000 долларов США.
- Серверная инфраструктура включает высокопроизводительные серверы с достаточным объёмом оперативной памяти и быстродействующими накопителями. Стоимость готовых конфигураций варьируется от 135 000 рублей в месяц для серверов с Tesla A100 до значительно более высоких сумм для промышленных решений.
- Сетевое оборудование обеспечивает высокоскоростную передачу данных между компонентами системы. Создание сетей передачи данных для 100 рабочих мест может потребовать инвестиций порядка 2.5 млн рублей.

2) Программные лицензии

Программное обеспечение составляет приблизительно 20 % от первоначальных затрат. Эта категория включает:

- Лицензии на платформы машинного обучения — стоимость корпоративных лицензий на такие платформы, как NVIDIA AI Enterprise, может достигать десятков тысяч долларов в год.

Например, бессрочные лицензии для ERP-системы на 20 пользователей оцениваются в 2.34 млн рублей.

- Операционные системы и СУБД — годовая стоимость лицензий на системы управления базами данных может составлять 36 000 рублей и более.
- Специализированное ПО для разработки и управления моделями включает инструменты для версионирования моделей, мониторинга производительности и автоматизации процессов MLOps.

3) Инфраструктура

Инфраструктурные затраты составляют около 25 % от общего бюджета и включают:

- Центры обработки данных — стоимость строительства ЦОД Tier 3 составляет 7 — 8 млн долларов США за 1 МВт для объектов мощностью свыше 20 МВт. Для дата-центров мощностью 5 — 20 МВт эта цифра возрастает до 8 — 9 млн долларов за 1 МВт. В российских условиях стоимость оснащения одного квадратного метра полезной площади может достигать 250 000 рублей.
- Системы охлаждения и электроснабжения — критически важные для обеспечения стабильной работы высокопроизводительного оборудования. Современные решения с непрямым испарительным охлаждением позволяют существенно снизить операционные затраты на электроэнергию.
- Системы безопасности и мониторинга обеспечивают физическую и информационную безопасность инфраструктуры.

4) Персонал

Затраты на персонал составляют приблизительно 15 % от первоначального бюджета. Команда проекта внедрения ГенИИ включает:

- Специалистов по машинному обучению — средняя зарплата ML-инженера в России составляет 160 000 рублей в месяц, с диапазоном от 90 000 до 230 000 рублей. В Москве средняя зарплата достигает 165 000 рублей в месяц.
- Архитекторов решений и DevOps-инженеров — обеспечивают проектирование и развёртывание инфраструктуры.
- Менеджеров проектов — координируют работу команды и взаимодействие с заказчиком. Стоимость руководителя проекта может составлять 300 000 рублей в месяц.

Ориентировочная ежемесячная стоимость команды для среднего бизнеса составляет от 600 000 до 2 000 000 рублей.

Матрица затрат по категориям и масштабам проектов

Матрица демонстрирует значительную вариативность затрат в зависимости от масштаба проекта. Стартапы могут ограничиться бюджетом в 6.5 млн рублей, в то время как крупные предприятия должны планировать инвестиции порядка 230 млн рублей для комплексного внедрения генеративных ИИ-систем [DailyMoscow-2024].

Категория затрат	Стартап (млн руб.)	Средний бизнес (млн руб.)	Предприятие (млн руб.)
Аппаратное обеспечение	2.5	15.0	80.0
Программные лицензии	1.0	8.0	35.0
Инфраструктура	1.5	12.0	60.0
Персонал (команда)	1.0	8.0	40.0
Интеграция/развёртывание	0.5	4.0	15.0
Итого	6.5	47.0	230.0

Методология расчёта и планирования

При планировании первоначальных затрат в соответствии с референтной моделью жизненного цикла рекомендуется использовать следующий поэтапный подход:

1. Формирование концепции — разработка концепции системы ГенИИ с включением проекта технического задания

2. Определение требований — анализ функциональных и технических требований к системе
3. Выбор архитектуры — принятие решения о модели развёртывания (облако или внутренний контур)
4. Расчёт ресурсов — определение необходимых вычислительных мощностей и объёмов данных
5. Планирование команды — формирование проектной команды с учётом компетенций и временных рамок
6. Оценка рисков — включение 10 — 20 % резерва на непредвиденные расходы

Особое внимание следует уделить этапу формирования концепции, где должен быть разработан проект технического задания, определяющий основные параметры будущей системы ГенИИ [Степанова, 2023].

1.4 Операционные расходы и их прогнозирование

Концептуальные основы операционных расходов ГенИИ-систем

Операционные расходы (ОРЕХ) генеративных систем искусственного интеллекта представляют собой совокупность текущих затрат, необходимых для поддержания функционирования системы на протяжении всего срока её эксплуатации. В отличие от капитальных затрат, операционные расходы носят периодический характер и прямо влияют на операционную прибыль организации [TBank-ОРЕХ].

Согласно российской практике, операционные расходы включают в себя все затраты, связанные с текущей деятельностью компании и влияющие на её операционную прибыль. Для ГенИИ-систем эти расходы имеют специфические особенности, обусловленные высокой вычислительной сложностью процессов обучения и функционирования моделей, необходимостью постоянного обновления и высокими требованиями к квалификации персонала.

Операционные затраты ГенИИ-систем характеризуются высокой волатильностью и зависимостью от интенсивности использования. Исследования показывают, что правильное планирование и управление операционными расходами может привести к сокращению затрат на 15 — 35 % при одновременном повышении производительности на 25 — 40 % [FedAG-ROI].

Структура операционных расходов

1) Энергопотребление

Энергопотребление является одной из наиболее существенных и постоянно растущих статей операционных расходов ГенИИ-систем. Современные GPU-серверы потребляют значительное количество электроэнергии: например, один GPU NVIDIA V100 потребляет до 300 Вт в час, что при стоимости 745.21 рубля за час работы составляет существенную статью расходов.

Энергопотребление GPU-систем характеризуется нелинейной зависимостью от частоты работы и нагрузки. Исследования показывают, что существует

оптимальная точка работы GPU, при которой достигается минимальное энергопотребление при максимальной производительности [Вечер и др., 2016]. При повышении частоты GPU сверх оптимального значения энергопотребление растёт быстрее, чем производительность.

Для расчёта энергозатрат рекомендуется использовать следующую формулу:

$$\begin{aligned} \text{Ежемесячные затраты на энергопотребление} = \\ \text{Мощность GPU (кВт)} \times \text{Часы работы в месяц} \times \text{Тариф за} \\ \text{кВт} \cdot \text{ч} \times \text{Коэффициент нагрузки} \end{aligned}$$

где коэффициент нагрузки учитывает фактическое использование вычислительных ресурсов и обычно составляет 0.6 — 0.8 для производственных систем.

2) Обслуживание и техническая поддержка

Техническое обслуживание ГенИИ-систем включает в себя мониторинг производительности, диагностику неполадок, профилактическое обслуживание оборудования и устранение технических проблем. Стоимость технической поддержки обычно составляет 15 — 20 % от общих операционных расходов.

Современные ГенИИ-системы требуют специализированного обслуживания, включающего:

- Мониторинг состояния моделей и качества их работы
- Контроль производительности вычислительной инфраструктуры

- Обеспечение безопасности данных и системы
- Управление версиями моделей и их развёртыванием

Системы машинного обучения требуют постоянного внимания MLOps-специалистов, так как модели нуждаются в регулярной проверке качества и перетренировке. Стоимость курсов MLOps для подготовки специалистов составляет от 54 000 до 190 000 рублей, что отражает высокую стоимость квалифицированного персонала.

3) Обновления и перетренировка моделей

Обновление моделей представляет собой одну из наиболее критических и дорогостоящих операций в жизненном цикле ГенИИ-систем. В отличие от традиционного программного обеспечения, модели машинного обучения требуют регулярной перетренировки для поддержания актуальности и точности работы.

Стоимость обучения современных генеративных моделей достигла беспрецедентного уровня. Для обучения GPT-4 потребовалось 78 млн долларов США, а для Google Gemini Ultra — 191 млн долларов США. Даже модели среднего размера требуют значительных затрат: обучение модели RoBERTa Large обошлось в 160 000 долларов США.

Прогнозы показывают, что если тенденция роста затрат сохранится, к 2027 году стоимость обучения передовых моделей ИИ превысит 1 млрд долларов США.

Для российского рынка стоимость использования генеративных моделей варьируется в зависимости от сложности. Например, использование YandexGPT Pro

стоит 1.20 рубля за 1000 токенов в синхронном режиме, что при интенсивном использовании может составить значительные суммы.

4) Поддержка и лицензирование

Поддержка программного обеспечения включает в себя лицензии на специализированные платформы, операционные системы и средства разработки.

Годовая стоимость лицензий может составлять от 36 000 рублей для базовых СУБД до сотен тысяч рублей для промышленных решений.

Для корпоративных решений стоимость может достигать значительных сумм: например, продвинутая модель OpenAI уровня PhD оценивается в 20 000 долларов США в месяц, что составляет около 2 млн рублей в год по текущему курсу.

Методология прогнозирования операционных расходов

Модели прогнозирования. Прогнозирование операционных расходов ГЕНИИ-систем требует применения специализированной методологии, учитывающей специфику этой области. Современные подходы включают использование собственно технологий искусственного интеллекта для прогнозирования затрат.

Базовая модель прогнозирования операционных расходов включает следующие компоненты:

$$\text{OPEX}(t) = \text{Энергозатраты}(t) + \text{Персонал}(t) + \text{Поддержка}(t) + \text{Обновления}(t) + \text{Лицензии}(t)$$

где t — временной период (месяц, квартал, год).

Каждый компонент может быть детализирован:

$$\text{Энергозатраты}(t) = \text{Базовое потребление} \times \text{Коэффициент загрузки}(t) \times \text{Тариф}(t)$$

$$\text{Персонал}(t) = \text{Количество сотрудников} \times \text{Средняя зарплата}(t) \times \text{Коэффициент индексации}$$

$$\text{Обновления}(t) = \text{Частота обновлений} \times \text{Стоимость переобучения} \times \text{Сложность модели}$$

Факторы, влияющие на динамику затрат. Операционные расходы ГенИИ-систем подвержены влиянию множества факторов:

- Масштабируемость нагрузки: увеличение объёма обрабатываемых данных приводит к пропорциональному росту энергозатрат и требований к вычислительным ресурсам.
- Технологическое развитие: появление новых, более эффективных архитектур может снизить операционные расходы, но требует инвестиций в обновление.
- Регуляторные требования: изменения в законодательстве могут потребовать дополнительных мер безопасности и соответствия.

Формулы расчёта ежемесячных затрат

Базовая формула расчёта ежемесячных операционных затрат:

Ежемесячные OPEX =

- + (Количество GPU × Стоимость часа × Часы работы в месяц)
- + (Количество специалистов × Средняя зарплата)
- + (Потребление электроэнергии × Тариф за кВт·ч)
- + (Ежемесячные лицензии и подписки)
- + (Амортизированная стоимость обновлений / 12)
- + (Расходы на техническую поддержку)

Примеры для различных масштабов проектов:

1) Стартап (до 10 пользователей):

- GPU: 1 — 2 карты Tesla T4, ~155 000 руб./мес. за GPU
- Персонал: 1 — 2 ML-инженера, ~320 000 руб./мес.
- Энергопотребление: ~15 000 руб./мес.
- Итого: ~650 000 руб./мес.

2) Средний бизнес (10 — 100 пользователей):

- GPU: 4 — 8 карт NVIDIA A100, ~1 600 000 руб./мес. за GPU
- Персонал: 3 — 5 специалистов, ~1 200 000 руб./мес.

- Энергопотребление: ~80 000 руб./мес.
- Поддержка и лицензии: ~400 000 руб./мес.
- Итого: ~3 500 000 руб./мес.

3) Крупное предприятие (100+ пользователей):

- GPU-кластер: 20+ карт высокого класса, ~8 000 000 руб./мес.
- Персонал: 10+ специалистов, ~3 500 000 руб./мес.
- Энергопотребление: ~300 000 руб./мес.
- Инфраструктура и поддержка: ~2 000 000 руб./мес.
- Итого: ~15 000 000 руб./мес.

Коэффициент операционных расходов (OER). Для анализа эффективности операционных затрат рекомендуется использовать коэффициент операционных расходов:

$$\text{OER} = (\text{Общие операционные расходы} / \text{Валовой доход}) \times 100\%$$

Оптимальное значение OER для ГенИИ-проектов составляет 60 — 75 % на этапе становления и 40 — 60 % для зрелых проектов.

Оптимизация операционных расходов

Для минимизации операционных расходов при сохранении эффективности ГенИИ-систем рекомендуется:

- Использование облачных решений с оплатой по факту использования для проектов с переменной нагрузкой
- Применение методов дистилляции моделей для снижения требований к вычислительным ресурсам
- Автоматизация процессов MLOps для сокращения трудозатрат
- Регулярный мониторинг и оптимизация энергопотребления
- Планирование обновлений моделей на основе метрик деградации качества

1.5 Скрытые затраты и риски

Концептуальные основы скрытых затрат ГенИИ-систем

Скрытые затраты при внедрении генеративных систем искусственного интеллекта представляют собой одну из наиболее недооценённых и опасных составляющих совокупной стоимости владения. Согласно международной практике, скрытые затраты составляют до 80 % от общих расходов на ИИ-проекты, в то время как организации традиционно фокусируются только на видимой части «айсберга» — прямых капитальных и операционных расходах [Reeng-2025].

Исследования показывают, что более 80 % проектов внедрения систем бизнес-аналитики выходят за рамки изначально запланированного бюджета именно из-за неучёта скрытых затрат [SiglaVision-BI]. Аналогичная ситуация наблюдается и при внедрении ГенИИ-систем, когда скрытые расходы могут в 3 — 4 раза превышать первоначальные планы.

По оценкам экспертов, скрытые ИТ-затраты составляют до 16 % от общих ИТ-затрат компании и часто оседают в бюджетах подразделений, в которых возникла проблема, что делает их практически невидимыми для финансового контроля [IEMag-Hidden-Costs]. В контексте ГенИИ-систем эта проблема усугубляется высокой сложностью технологии и необходимостью глубокой интеграции с существующими бизнес-процессами.

Феномен скрытых затрат в ИТ-проектах объясняется тем, что при планировании кампании традиционно концентрируются на очевидных статьях расходов — лицензиях на программное обеспечение и затратах на оборудование. Однако это лишь «верхушка айсберга». За каждый рубль, потраченный на лицензии, компании тратят еще как минимум 1 — 3 рубля на сопутствующие расходы.

Затраты на генеративный ИИ

**Видимые
затраты
20%**



Структура скрытых затрат

1) Обучение персонала и адаптация команды

Обучение персонала представляет собой одну из наиболее существенных статей скрытых расходов при внедрении ГенИИ-систем. Современные генеративные технологии требуют кардинально нового подхода к работе с данными и принятию решений, что требует существенных усилий по обучению имеющегося персонала.

Стоимость специализированных курсов MLOps составляет от 54 000 до 190 000 рублей за одного специалиста. При этом для эффективного внедрения ГенИИ-системы требуется обучение не только технических специалистов, но и конечных пользователей, менеджеров среднего звена и руководителей высшего уровня.

Исследования показывают, что компании часто недооценивают время, необходимое для адаптации персонала к новым технологиям. Процесс освоения ГенИИ-инструментов может занимать от 3 до 12 месяцев в зависимости от сложности внедряемой системы и квалификации сотрудников [Computerra-2025]. В течение этого периода производительность может снижаться на 15 — 30 %, что представляет собой значительные косвенные потери.

2) Интеграция с существующими системами

Интеграция ГенИИ-систем с существующей ИТ-инфраструктурой часто оказывается гораздо более сложной и дорогостоящей задачей, чем изначально планируется. Проблемы возникают на уровне

совместимости форматов данных, API различных систем, требований к безопасности и производительности.

Недостаточное тестирование интеграционных решений может привести к критическим проблемам с совместимостью, потере данных и нарушению отлаженных бизнес-процессов. На этом этапе дополнительные траты возникают из-за простоев, необходимости срочно устранять неполадки и восстанавливать базу данных [TU-Dop-2023].

Особую сложность представляет интеграция с унаследованными системами, которые могут не поддерживать современные API или требовать значительной модернизации для совместной работы с ГенИИ-решениями. Стоимость такой модернизации может составлять 50 — 100 % от стоимости самой ГенИИ-системы.

3) Простои и потери производительности

Простои при внедрении ГенИИ-систем представляют собой одну из наиболее критических категорий скрытых затрат. Даже при самом тщательном планировании переход на новые технологии неизбежно связан с временными нарушениями рабочих процессов.

Исследования показывают, что в первые 6 месяцев после внедрения ИИ-систем производительность сотрудников может снижаться на 20 — 40 % из-за необходимости адаптации к новым инструментам и процессам [FedAG-ROI]. Для крупных организаций такое снижение производительности может обойтись в млн рублей упущенной выгоды.

4) Миграция и очистка данных

Качество данных является критическим фактором успеха ГенИИ-проектов. Однако подготовка данных для обучения генеративных моделей часто требует значительно больше времени и ресурсов, чем изначально планируется.

Процесс очистки и структурирования данных может занимать 60 — 80 % от общего времени проекта внедрения ИИ. Это включает в себя выявление и устранение дубликатов, нормализацию форматов, исправление ошибок, обеспечение консистентности данных и их аннотацию для обучения моделей [CBM-Solutions].

Миграция данных из различных источников в единую платформу также представляет собой сложную задачу, особенно при работе с большими объёмами информации. Ошибки на этом этапе могут привести к необходимости повторного обучения моделей, что значительно увеличивает сроки и стоимость проекта.

5) Изменение бизнес-процессов

Внедрение ГенИИ-систем часто требует фундаментального переосмысления существующих бизнес-процессов. Это связано не только с техническими возможностями новых систем, но и с изменением ролей сотрудников, принципов принятия решений и методов контроля качества.

Реорганизация бизнес-процессов может потребовать:

- Пересмотра должностных инструкций и регламентов
- Изменения систем мотивации и KPI
- Обновления процедур контроля качества

- Адаптации систем отчётности

Эти изменения часто встречают сопротивление со стороны персонала, что может потребовать дополнительных инвестиций в управление изменениями и коммуникационные программы.

6) Управление изменениями

Управление изменениями представляет собой критически важный аспект успешного внедрения ГЕНИИ-систем, составляющий около 10 % от общих скрытых затрат. Стоимость профессиональных программ управления изменениями варьируется от 70 до 134 тысяч рублей для группы 10 — 14 сотрудников [Games4Business].

Сопротивление сотрудников инновациям является одной из ключевых проблем внедрения новых технологий [Listohod-SED]. Люди, привыкшие к традиционным процессам работы, часто испытывают дискомфорт перед необходимостью осваивать новые технологии. Это может проявляться в форме активного протеста, пассивного саботажа или просто нежелания использовать новую систему в полную силу.

Основные компоненты затрат на управление изменениями включают:

- Обучение руководителей методам управления изменениями
- Разработку коммуникационной стратегии
- Создание мотивационных программ для сотрудников
- Психологическую поддержку персонала в период адаптации

7) Дополнительные риски и непредвиденные расходы

К прочим рискам, составляющим около 5 % от скрытых затрат, относятся:

- Технологические риски:

- Устаревание технологий и необходимость обновления
- Зависимость от поставщиков решений
- Проблемы масштабирования пилотных проектов
- Сложности интеграции с облачными сервисами

- Организационные вызовы:

- Недостаток квалифицированных специалистов по ИИ
- Необходимость реинжиниринга бизнес-процессов
- Проблемы с принятием решений на основе рекомендаций ИИ

- Правовые и регуляторные риски:

- Изменения в законодательстве о персональных данных
- Требования к объяснимости решений ИИ
- Необходимость соответствия отраслевым стандартам

Методология оценки и управления скрытыми затратами

Для корректной оценки скрытых затрат рекомендуется использовать следующую методологию:

1) Этап планирования:

- Проведение аудита готовности организации к внедрению ИИ
- Оценка квалификации персонала и потребностей в обучении

- Анализ существующей ИТ-архитектуры и требований к интеграции
- Разработка плана управления изменениями

2) Этап внедрения:

- Постоянный мониторинг производительности системы
- Контроль качества выходных данных ИИ-модели
- Измерение фактических затрат времени на адаптацию персонала
- Отслеживание простоев и их причин

3) Стратегии минимизации:

- Поэтапное внедрение с пилотными проектами
- Привлечение опытных консультантов на критических этапах
- Инвестирование в обучение ключевых сотрудников до начала внедрения
- Создание резервов на непредвиденные расходы в размере 20 — 30 % от основного бюджета

Успешное управление скрытыми затратами требует комплексного подхода, учитывающего не только технические аспекты внедрения, но и человеческий фактор, организационную культуру и долгосрочные стратегические цели компании.

Представленная модель помогает руководителям понять реальный масштаб инвестиций, необходимых для успешного внедрения ГенИИ-систем, и заранее планировать соответствующие резервы.

Коэффициенты увеличения бюджета. Для защиты от превышения бюджета рекомендуется использовать систему коэффициентов, учитывающих различные факторы риска:

- Если проект согласуют два и более лица — бюджет увеличивается в 1.5 раза
- Если заказчик не ориентируется в предметной области — коэффициент 1.2
- При работе с унаследованными системами — коэффициент 1.3 — 1.5
- При отсутствии опыта внедрения ИИ в организации — коэффициент 1.4

Планирование резервов на непредвиденные расходы. Согласно практике строительной отрасли, резерв на непредвиденные расходы должен составлять 10 — 15 % от общей стоимости проекта. Для ГенИИ-проектов, учитывая их инновационный характер и высокую степень неопределённости, рекомендуется увеличить этот резерв до 20 — 25 %.

Резерв должен покрывать:

- Превышение сроков разработки
- Дополнительные требования к вычислительным ресурсам
- Непредвиденные проблемы с интеграцией
- Необходимость дополнительного обучения персонала
- Изменения в регуляторных требованиях

Поэтапный подход к внедрению ГенИИ-систем позволяет выявлять и устранять проблемы на ранних стадиях, снижая общие риски проекта.

Рекомендуется начинать с пилотных проектов ограниченного масштаба, постепенно расширяя функциональность и охват.

Этапы внедрения должны включать:

- Анализ готовности организации
- Пилотный проект (3-6 месяцев)
- Анализ результатов и корректировка подхода
- Поэтапное масштабирование

Полное внедрение с мониторингом эффективности
Тщательное планирование и анализ рисков. Детальный анализ рисков на этапе планирования помогает выявить потенциальные источники скрытых затрат и заранее подготовить стратегии их митигации. Необходимо проводить оценку технических, организационных, финансовых и регуляторных рисков.

Привлечение опытных интеграторов.
Сотрудничество с опытными интеграторами, имеющими успешный опыт внедрения ГенИИ-систем, позволяет избежать типичных ошибок и существенно снизить скрытые затраты. Стоимость качественных консультационных услуг обычно окупается за счёт экономии на исправлении ошибок.

Понимание и правильное планирование скрытых затрат является критически важным фактором успеха проектов внедрения ГенИИ-систем. Организации, которые недооценивают эти расходы, рискуют не только превысить бюджет, но и поставить под угрозу достижение поставленных целей автоматизации и цифровой трансформации.

2 Классификация архитектурных решений

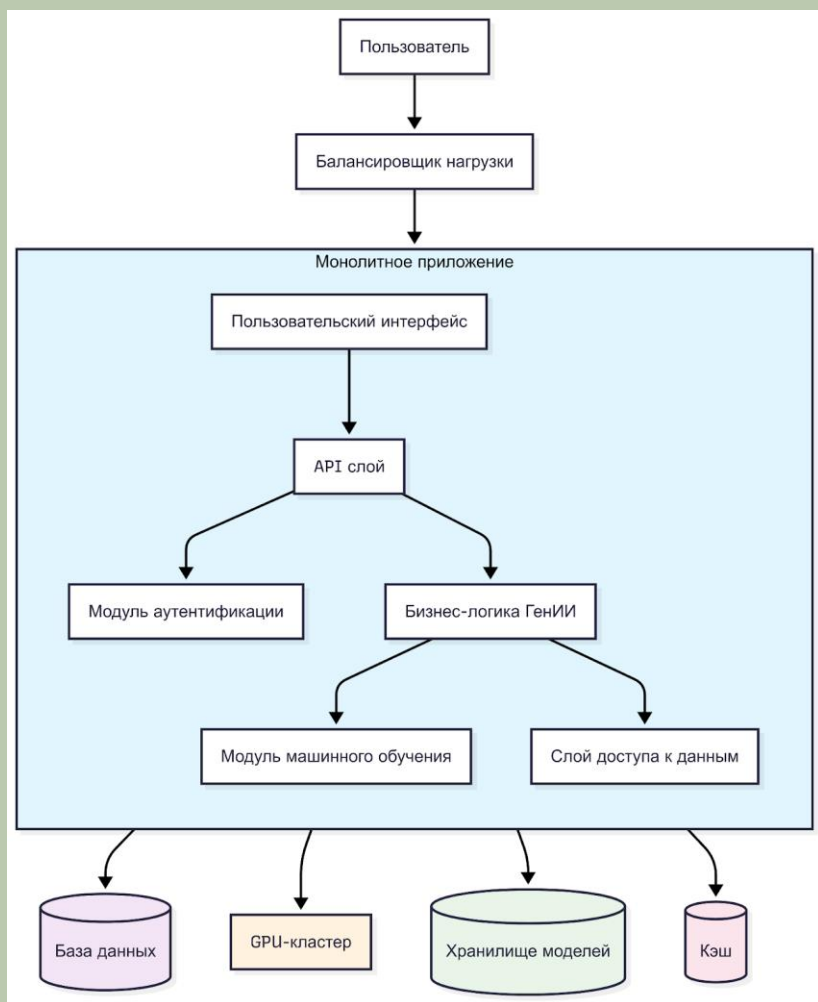


2.1 Монолитная архитектура

Единственный исполняемый модуль содержит пользовательский интерфейс, бизнес-логику и доступ к данным.

Плюсы — простота развёртывания, минимальные накладные расходы, единый контур наблюдаемости.

Минусы — трудно масштабировать отдельные функции, общий цикл выпуска, ограниченная технологическая гибкость.



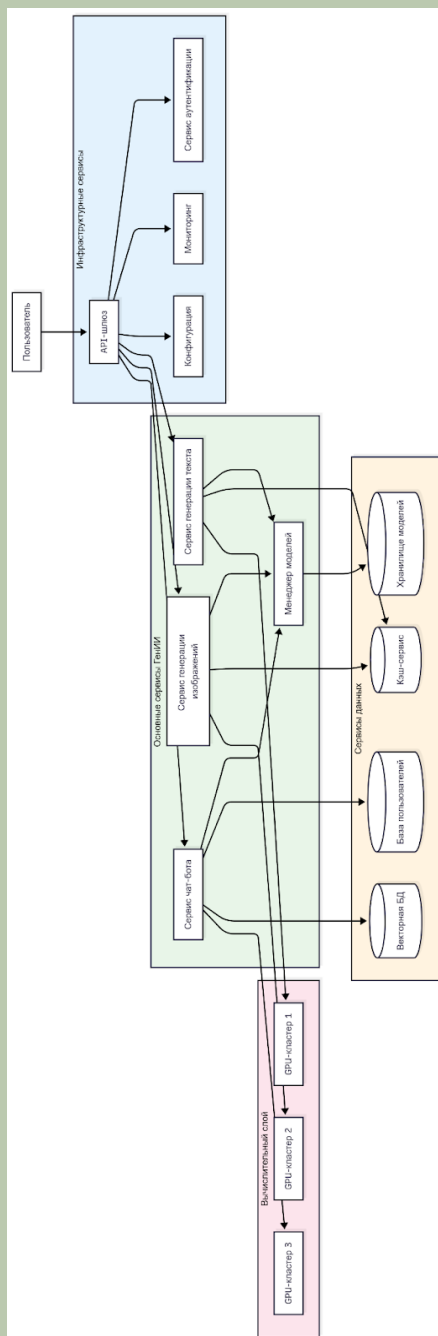
Все слои генерируют единый контейнер/процесс.

2.2 Микросервисная архитектура

Приложение разбивается на изолированные службы, каждая решает одну бизнес-функцию и общается по легковесному протоколу (обычно REST/gRPC) [Кокотов, 2023].

Плюсы — независимые выпуски, горизонтальное масштабирование, технологический плюрализм.

Минусы — сложность оркестрации, межсервисная согласованность данных, необходимость стека для мониторинга состояния системы.

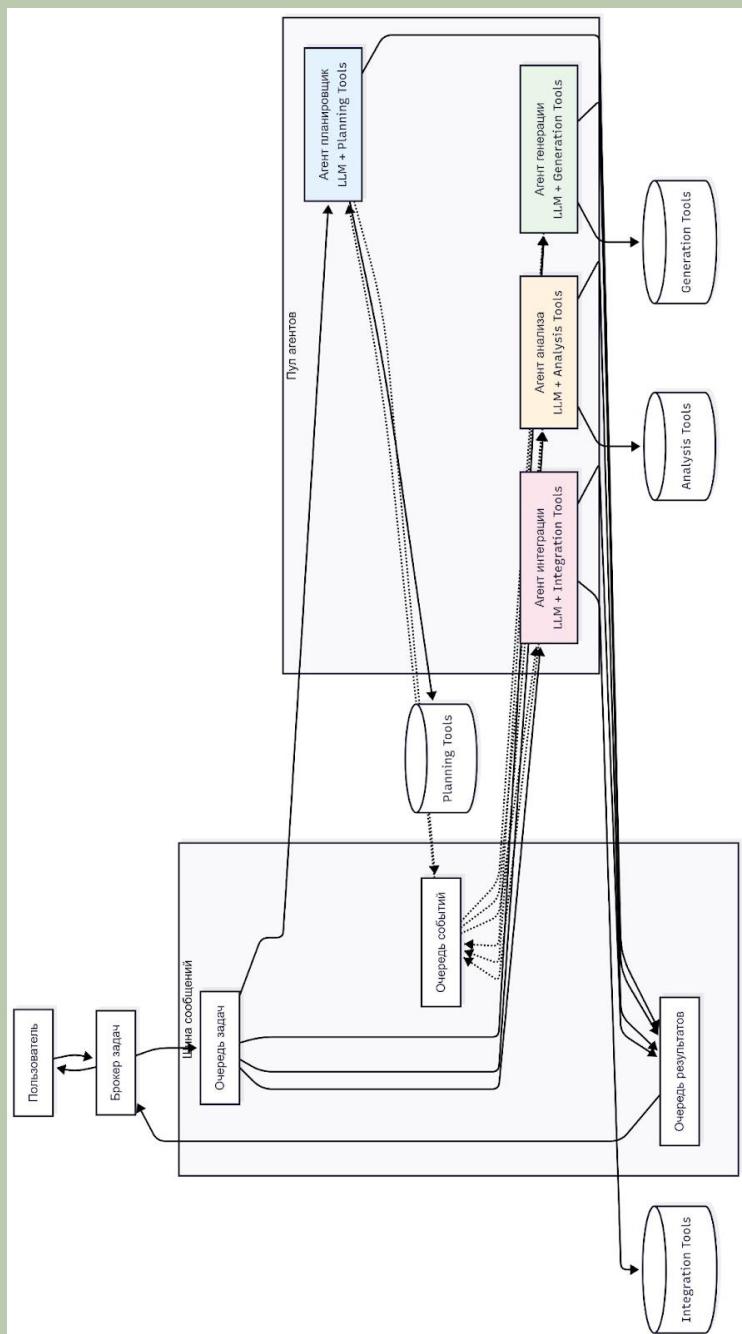


2.3 Многоагентные системы

Система строится из автономных агентов — рабочих процессов БЯМ или специализированных сервисов, которые координируют действия через брокер задач либо шину событий.

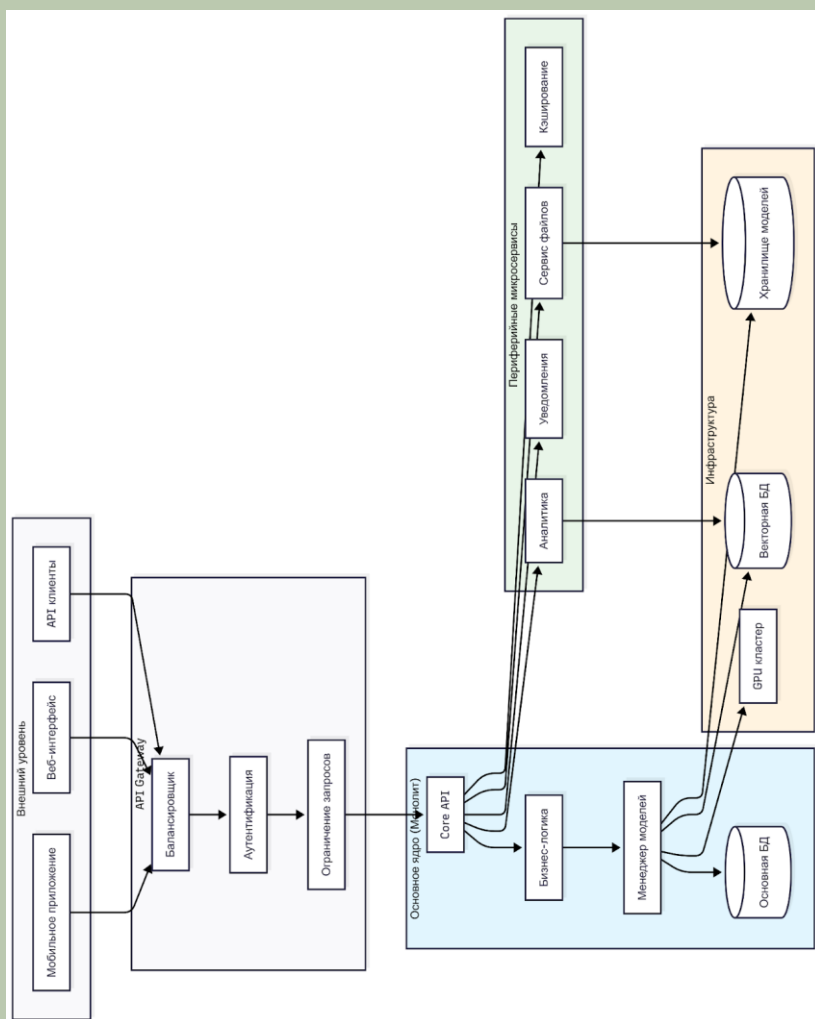
Плюсы — устойчивость к отказам, естественное масштабирование по потребности рабочего процесса.

Минусы — сложность планирования и проверки корректности коллективного поведения, рост сетевых задержек.



2.4 Гибридные модели

Сочетают преимущества монолита (низкая латентность внутри домена) и микросервисов (изолируемость). Типовой подход — «внутри сервиса монолит, между сервисами API» или «ядро-монолит + периферия-микросервисы».



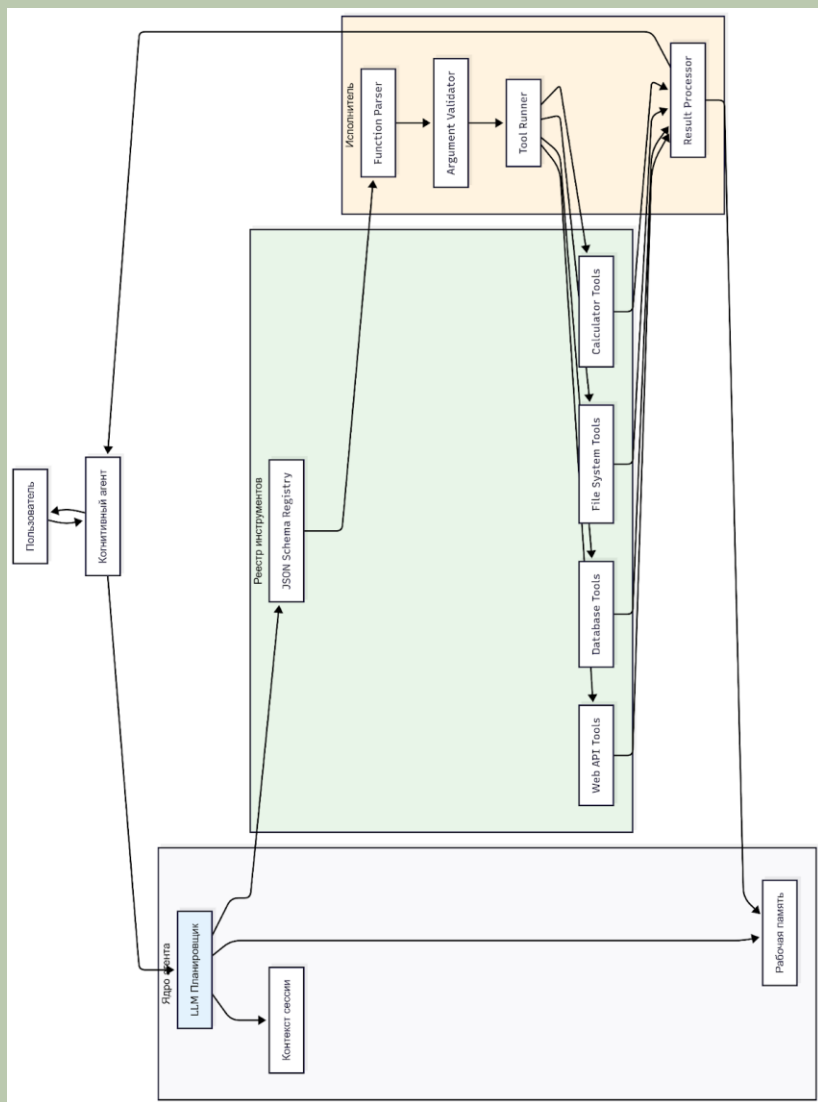
2.6 БЯМ как ядро когнитивного агента

Современные модели (GPT-4o, CoALM-70B) получают расширенный «репертуар» функций через механизм вызова функций: модель формирует JSON-структуру для вызова внешнего API, агент исполняет, возвращает результат модели.

Инструментальная шина — реестр функций JSON Schema;

БЯМ-планировщик определяет, какой инструмент применить;

Исполнитель запускает инструмент по протоколу HTTP/CLI и передаёт вывод назад в контекст БЯМ.



2.7 Сравнительная таблица

Тип архитектуры	Масштабирование	Сложность DevOps	Задержка	Подходит для
Монолит	Вертикальное	Низкая	Минимальная	Прототипы, MVP
Микросервисы	Горизонтальное (по сервисам)	Средняя/высокая	Средняя	B2C-порталы, высоконагруженные API
Многоагентная	Горизонтальное (по агентам)	Высокая	Средняя/высокая	Автономные рабочие процессы, RPA
Гибридная	Комбинированное	Средняя	Низкая/средняя	Корпоративные ядра + мобильные фронты
RAG-стек	Горизонтальное (вывод/поиск)	Средняя	Средняя	Справочные боты, поиск знаний
БЯМ + инструменты	Горизонтальное (БЯМ/инструменты)	Средняя	Средняя	Агенты-ассистенты, автоматизация рабочих процессов

3 Сравнительный анализ облачных и локальных решений



3.1 Модели развёртывания и их характеристики

Переход к генеративным ИИ-системам возможен четырьмя базовыми способами размещения вычислительных ресурсов:

- 1) Публичное облако. Инфраструктура полностью принадлежит провайдеру; потребление ресурсов осуществляется по модели pay-as-you-go. Ключевые преимущества — нулевой CAPEX, мгновенная масштабируемость и встроенные механизмы георезервирования [SelectelPublic, AzureReliability]. Ограничения — зависимость от провайдера, необходимость тщательного контроля комплаенса и риски локализации данных [ComplianceCloud, TrendMicro].
- 2) Частное облако. Физические серверы и средства виртуализации выделяются одной организации. Решение предпочтительно для отраслей с повышенными требованиями к конфиденциальности данных (финансы, госзаказ) [ITGradPrivate]. CAPEX и расходы на обслуживание выше, но достигается полный административный контроль и предсказуемая производительность.
- 3) Гибридное решение. Объединяет публичное и частное облако с локальным ЦОД через защищённые каналы; позволяет выносить пиковые или низкокритичные нагрузки в общественное облако, сохраняя чувствительные данные на собственной инфраструктуре [FSHybrid, YandexSec]. Главные вызовы — усложнение интеграции, единое управление доступом и мониторинг.

- 4) Локальная инфраструктура (on-premise). Классическая модель собственных серверов в корпоративном ЦОД. Обеспечивает минимальную сетевую задержку и полную автономию, однако требует крупных инвестиций в оборудование, энерго- и охлаждающие мощности, штат администраторов, а также регулярного обновления парка серверов [Попов, 2020].

Сравнительная матрица моделей развёртывания

Критерий	Публичное облако	Частное облако	Гибрид	On-premise
CAPEX	Нулевой	Высокий — закупка оборудования	Средний — частичный перенос в облако	Максимальный — полный парк серверов
OPEX-предсказуемость	Плата по факту, зависит от трафика	Стабильные, но выше из-за обслуживания	Смешанные расходы	Высокие, но управляемые внутри компании
Масштабируемость	Неограниченная через портал	Ограничена ресурсами ЦОД	Эластичность публичной части	Требует закупки нового «железа»
Время развертывания	Минуты	Недели (поставка серверов)	Минуты / недели	Недели–месяцы
Административный контроль	Низкий	Полный	Дифференцированный	Полный
Безопасность периметра	Модель Zero Trust провайдера	Выделенный сегмент сети	Объединённые политики	Традиционный периметр
Соответствие требованиям	Требует договоров, SLA и сертификатов	Проще обеспечить изоляцию	Надо синхронизировать две среды	Полная внутренняя юрисдикция

Критерий	Публичное облако	Частное облако	Гибрид	On-premise
Отказоустойчивость	Зоны доступности, DR-сервисы	Зависит от дизайна частного ЦОД	Высокая при правильной маршрутизации	Требуется дублирование площадок
Латентность	Зависит от канала к провайдеру	Минимальна внутри сети	Минимальна для локальных сервисов	Минимальна
Предсказуемость производительности	Может колебаться при «шумных соседях»	Гарантирована выделенными ресурсами	Зависит от участка нагрузки	Гарантирована
Интеграция с «наследием»	Через API / VPN, возможны сложности	Нативная	Требуется шлюзов и брокеров	Нативная
Фиксация вендоров	Высокая при использовании PaaS	Низкая, если используется on-premise source	Средняя	Низкая
Восстановление после отказов	Встроено (мульти-регион)	Нужно проектировать	Использует DR облака	Требуется собственной площадки
Энергоэффективность	Ответственность провайдера, оптимизированные ЦОД	Зависит от энергоэффективности оборудования	Делится между сторонами	Требуется капитальных вложений
3-летний TCO (относит.)	Низкий при переменной нагрузке	Средний / высокий	Средний	Высокий

Выводы:

- Публичное облако оптимально для проектов с переменной или трудно прогнозируемой нагрузкой и ограниченными начальными бюджетами.
- Частное облако выбирают организации с жёсткими регуляторными ограничениями или требующие полной изоляции ресурсов.
- Гибридная модель предоставляет баланс гибкости и контроля, но требует зрелых процессов безопасности и мониторинга.
- On-premise остаётся обоснованным там, где критичны минимальная задержка, полный контроль над данными и невозможность передачи их внешнему провайдеру.

3.2 Экономический анализ моделей развёртывания

Точки безубыточности

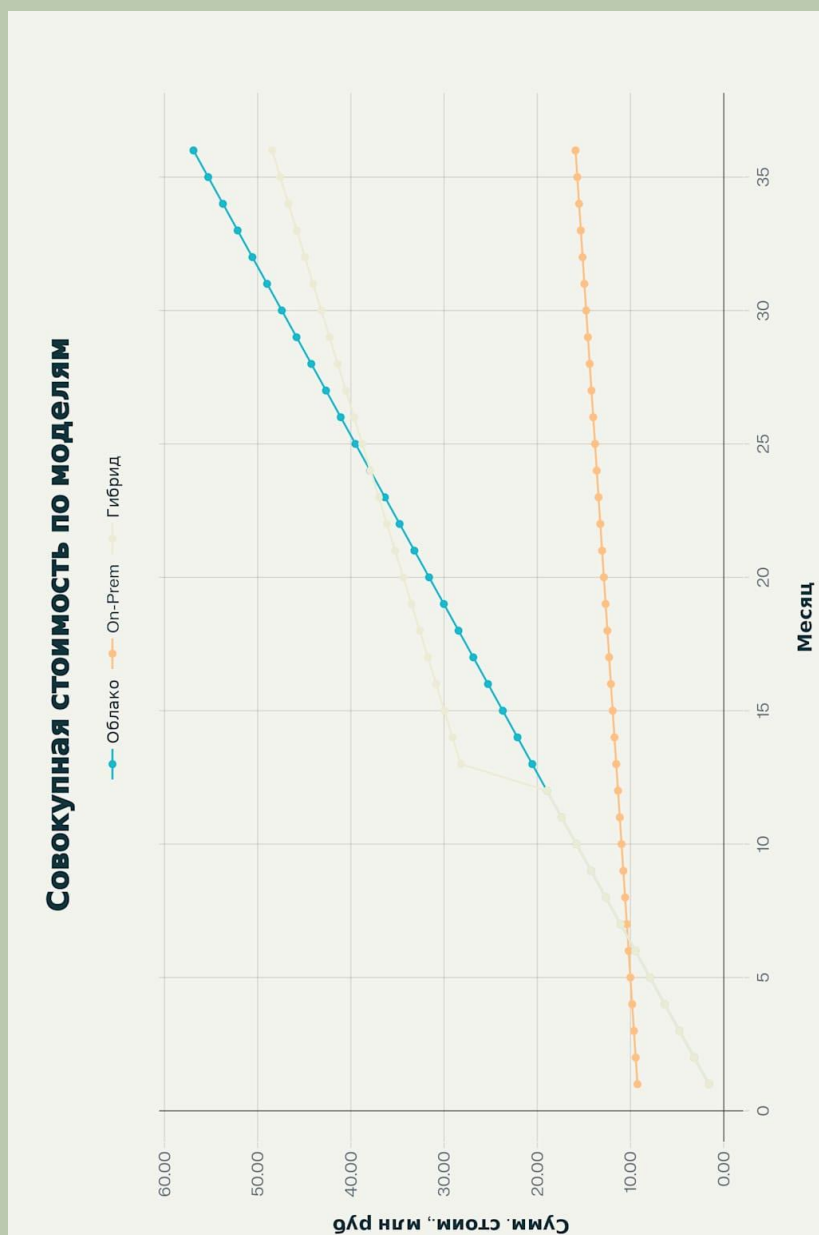
Модель	Первоначальные вложения	Средние OPEX, Р/мес	Кумулятивная ССВ = CAPEX + $\sum$ OPEX	Точка безубыточности, мес
Облако (на примере Yandex Cloud g2.4 — 4×A100)	0	1 580 000	1 580 000 × t	—
On-Prem (сервер 4×H100 + стойко-место)	9 055 000	190 000	9 055 000 + 190 000 × t	6
Гибрид (12 мес облако → покупка сервера)	0 до M12, затем 9 055 000 + OPEX 190 000	см. график	окупаемость ≈ 27 мес после старта	

Расчёты:

- Облако: $4 \text{ GPU A100} = 2\,164.8 \text{ Р/ч} \rightarrow 1\,580\,000 \text{ Р/мес}$ (730 ч).
- Оп-Prem: CAPEX = 9 055 000 Р (4 × H100-сервер) + монтаж; OPEX = 145 000 Р colocation + 13 000 Р электроэнергия ($3 \text{ кВт} \times 6 \text{ Р/кВт}\cdot\text{ч}$) + 32 000 Р поддержка.
- Гибрид: 12 мес в облаке, затем покупка оборудования.

Долгосрочный прогноз затрат

Кумулятивные потоки за 36 месяцев показывают, что после 6-го месяца Оп-Prem становится дешевле облака; гибрид обгоняет облако на 13-м месяце, но отыгрывает CAPEX лишь к 27-му месяцу.



Сравнение совокупной стоимости владения (ССВ) для облачной, локальной и гибридной моделей за 36 месяцев

Критерии выбора оптимальной модели

Критерий	Облако	On-Prem	Гибрид
Скорость запуска	Максимальная (минуты)	4–6 нед	Минуты → 4–6 нед
Капитальные затраты	0	Высокие	0 → Высокие
Прогнозируемость OPEX	Средняя (pay-as-you-go)	Высокая	Средняя
Масштабирование	Эластичное	Ограничено мощностью ЦОД	Комбинированное
Контроль над данными	Умеренный	Полный	Полный после миграции
Риски дефицита мощностей	Нет	Возможны (стойко-место 130–150 тыс ₽/мес)	Умеренные
Точка безубыточности	отсутствует	≤ 6 мес	≈ 27 мес
Рекомендуется для	MVP, переменные нагрузки	Долгие стаб.-нагрузки (≥ 3 года)	Быстрый старт → стабильная эксплуатация

Выводы:

- Для пилотных проектов и переменных нагрузок оптимальна облачная модель.
- Если нагрузка стабильна ≥ 3 лет и объём GPU-часов превышает $\sim 10\,000$ ч/год, выгоднее сразу переходить на собственное оборудование.
- Гибридный сценарий целесообразен, когда проект должен стартовать мгновенно, но через год-полтора ожидается стабильная высокая загрузка и доступна единовременная инвестиция в CAPEX.

3.3 Критерии выбора модели развёртывания

1) Технические критерии:

- Нагрузка и её волатильность. Переменные, пиковые потоки данных оправдывают использование публичного облака — эластичное масштабирование исключает простои или недозагрузку оборудования [ServerSpace-Checklist].
- Требуемая латентность. Задачи реального времени (≤ 20 мс) проще обеспечить в локальном ЦОДе; гибридная схема даёт возможность вынести критичные с точки зрения латентности сервисы «на землю», оставив фоновые задачи в облаке.
- Сложность технологического стека. Микросервисные и многоагентные решения требуют зрелой платформы оркестрации (K8s + наблюдаемость). Если в штате нет DevOps/MLOps-команды, предпочтительно облако с управляемыми сервисами [ITmatika-Platform].

- Надёжность и отказоустойчивость. Гарантированное SLA ≥ 99.95 % чаще предоставляет крупный провайдер; для on-prem это потребует резервного ЦОДа или договора с площадкой размещения оборудования.

2) Экономические критерии:

- CAPEX vs OPEX. При горизонте использования < 18 мес суммарные издержки облака обычно ниже из-за отсутствия крупных начальных вложений [ServerSpace-Checklist].
- Порог безубыточности. Для генеративных ИИ-кластеров «4 × H100» бьёт облачный g2.4 после ~6 месяцев непрерывной загрузки (см. § 3.2).
- Кэш-флоу проекта. Облачная модель сглаживает платежи, что особенно важно при бюджетировании госконтрактов с квартальной разбивкой.
- Финансовые риски. Рост тарифов на энергию и лицензии ложится на владельца on-prem; в облаке эти риски частично распределены между клиентами провайдера.

3) Правовые и комплаенс-критерии:

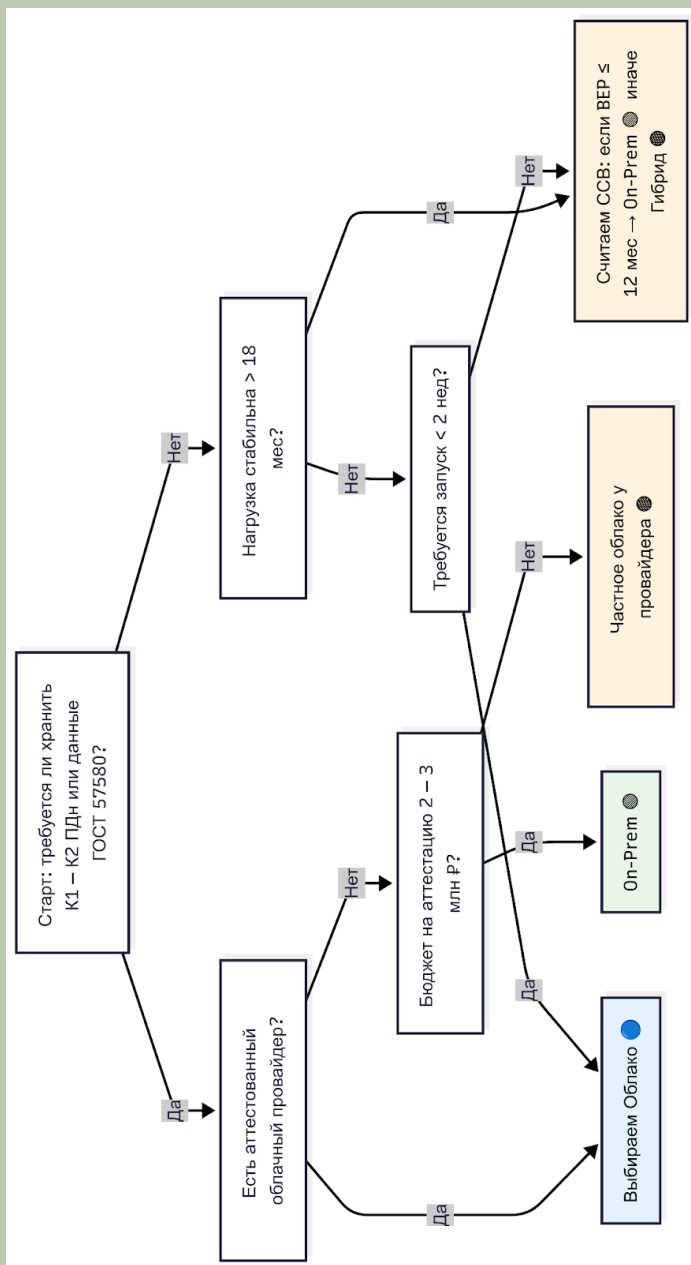
- Классы данных. Персональные данные уровня K1 — K2 по 152-ФЗ разрешены к размещению только в аттестованных облаках (ФСТЭК / ФСБ) [VKCloud-152, CloudMTS-GOST].
- Финансовая информация. Банковские ИС требуют соблюдения ГОСТ Р 57580.1—2017; сертифицированные площадки (K2 Cloud, Nubes, T1 Cloud) сокращают издержки на собственную

аттестацию [K2Cloud-GOST, Nubes-GOST, T1Cloud-GOST].

- Международные нормы. При трансграничных операциях учитываются GDPR, LGPD, CCPA; провайдер должен предоставлять договор-SCC или аналогичные механизмы [Runet-Compliance, TrendMicro].
- Лицензионные ограничения ПО / моделей (например, LLaMA 2) могут накладывать запрет на публичные сервисы, требуя on-prem или «изолированное» облако.

4) Организационные критерии:

- Скорость запуска. Пилоты и MVP обычно стартуют в облаке (минуты — часы), тогда как закупка железа и подготовка площадки занимают 4 — 8 недель.
- Квалификация команды. Недостаток экспертизы в управлении GPU-фермами — веский довод в пользу облака [AI-Success-CNews].
- Управление изменениями. Для организаций с жёсткими регламентами ИБ и управления изменениями проще использовать on-prem или частное облако в контуре заказчика.
- Культура DevOps / MLOps. Наличие зрелых процессов CI/CD, IaC и наблюдаемости упрощает переход к гибридной модели.



Дерево принятия решений по выбору схемы размещения

3.4 Миграционные стратегии

Грамотно спланированная миграция минимизирует простои, снижает риски снижения качества услуг и позволяет постепенно выйти на оптимальную модель развёртывания (облако / on-prem / гибрид). Ниже приведены основные сценарии перехода, типовые этапы и подходы к управлению рисками.

Этапы миграции

- Оценка (Assessment) — аудит нагрузки, данных, лицензий, ИБ-требований.
- Планирование (Planning) — выбор сценария, расчёт бюджетов, определение KPI и точек проверки.
- Пилот (Proof-of-Concept) — перенос ограниченного набора сервисов для проверки гипотез.
- Основной перенос (Execution) — поэтапный или одномоментный cut-over.
- Стабилизация и оптимизация (Stabilization) — тонкая настройка, оптимизация затрат, вывод устаревшей инфраструктуры из эксплуатации.

Сценарии перехода между моделями

Сценарий	Когда применяют	Ключевые особенности	Основные риски
Big Bang («одномоментный переход»)	Простые окружения, жёсткие сроки	Одновременное переключение всех сервисов	Увеличенный шанс отказа; ограниченные возможности отката [ITGLOBAL- VMwareCloud]
Phased Migration (поэтапная миграция)	Большие или гетерогенные системы	Перенос по модулям/бизнес- процессам с параллельной синхронизацией	Требует сложной координации и двойной поддержки [Kominform- DataMig]
Parallel Run / Blue-Green	Критичные сервисы, нужна проверка в боевых условиях	Одновременная работа старой и новой среды; трафик постепенно переводится	Рост OPEX на период параллельной работы
Hybrid Bridge	Долгие проекты, частичный перенос	«Мост» связывает облачные и локальные ресурсы; данные постепенно реплицируются	Увеличенная архитектурная сложность

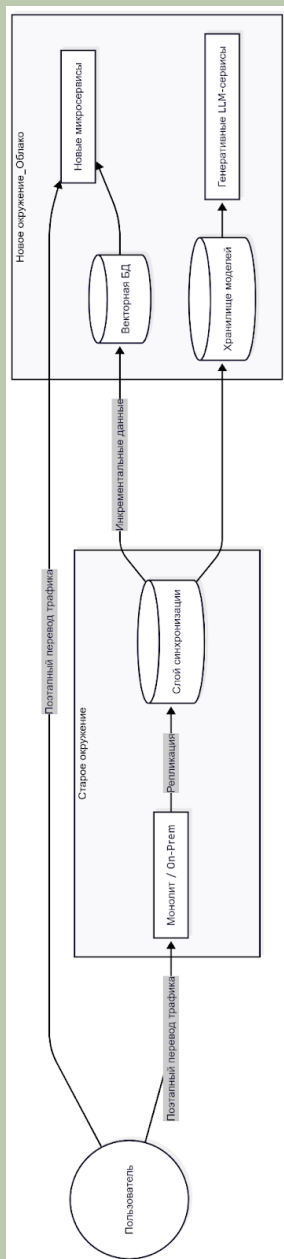
Управление рисками

Категория риска	Мероприятия по снижению
Технический — несовместимость ПО, сбой в cut-over	Двойное окружение, автоматические тесты, планы по откатам [ServerSpace-Checklist]
Данные — потеря / несогласованность	Синхронная репликация, контроль контрольных сумм, пилот на подмножестве данных [Komiinform-DataMig]
Производительность	Нагрузочное тестирование, SLA с провайдерами [ITmatika-Platform]
Организационный — нехватка компетенций	Обучение команды, привлечение подрядчиков, чистая RACI-матрица [AI-Success-CNews]
Финансовый — превышение бюджета	Резерв 15 — 25 % на непредвиденные расходы, контроль CAPEX/OPEX по фазам проекта

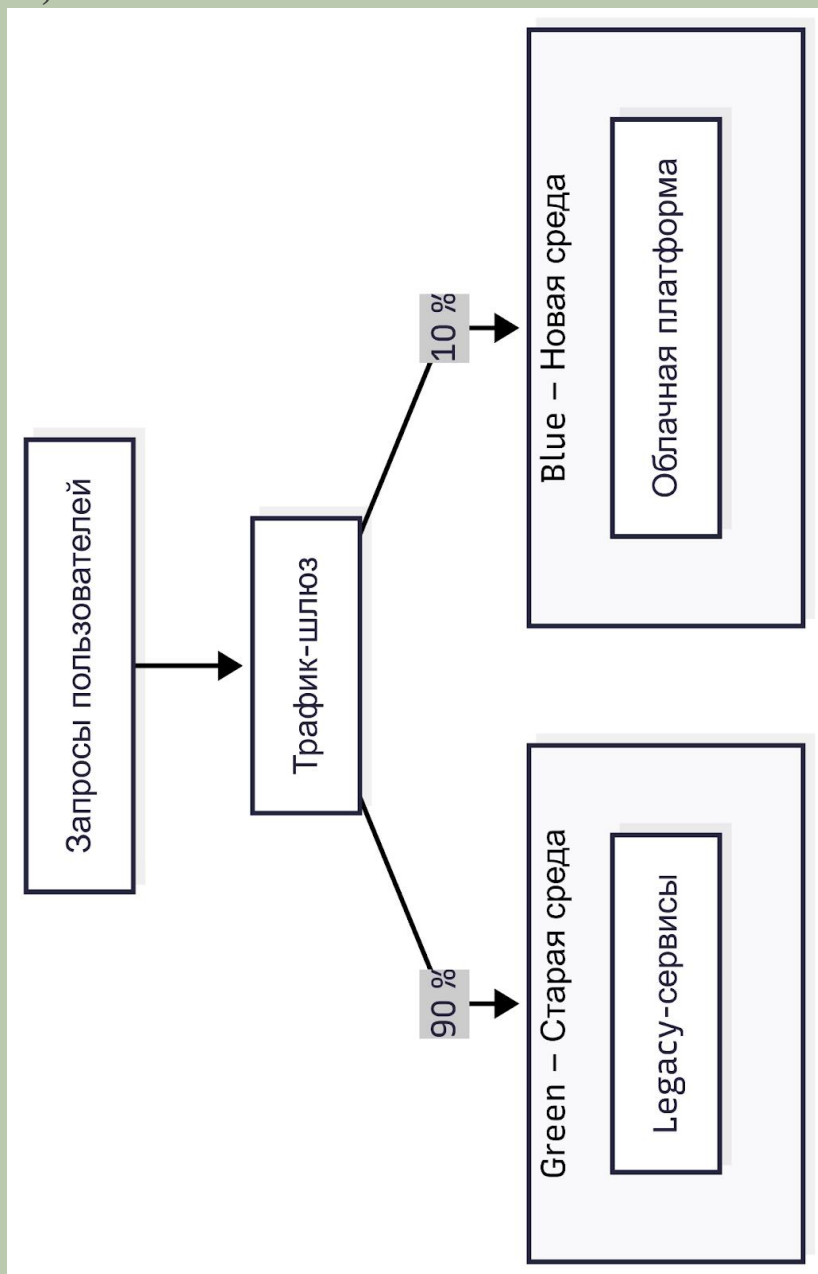
1) Временной план миграции



2) Phased Migration (частичная синхронизация)



3) Blue-Green / Parallel Run



4 Рекомендации для заказчиков



4.1 Методология планирования проекта внедрения

Планирование проекта внедрения системы ГенИИ должно соответствовать референтной модели жизненного цикла, описанной в подразделе 1.1. Каждый этап планирования соотносится с соответствующими стадиями жизненного цикла:

- Этап инициирования → Стадия формирования требований
- Этап проектирования → Стадия разработки концепции
- Этап PoC → Стадия технического проектирования
- Этап пилотной эксплуатации → Стадия рабочего проектирования
- Этап масштабирования → Стадия ввода в действие
- Этапы адаптации и поддержки → Стадия сопровождения
- Такое соответствие обеспечивает системный подход к управлению проектом и соблюдение требований российских стандартов.

1) Этап 1. Инициирование

- Формулировка бизнес-целей и KPI (рост выручки, % экономии затрат) [LinkedIn-AI-Roadmap].
- Аудит данных, ИТ-ландшафта и регуляторных ограничений.
- Ключевое решение — целесообразность ГенИИ и предельный бюджет.

2) Этап 2. Проектирование

- Приоритизация вариантов использования (матрица «ценность/сложность»).
- Выбор целевой архитектуры (облако, on-prem, гибрид), модели безопасности и ответственной роли владельца продукта [MS-Azure-ML].
- Разработка финансового плана (CAPEX / OPEX) и календаря работ.

3) Этап 3. Proof-of-Concept (PoC)

- Сбор пилотного датасета, настройка минимальной инфраструктуры.
- Метрики успеха: точность модели, Latency < 300 мс, TCO $\leq$ порога.
- Решение — «идти дальше / переработать / остановить».

4) Этап 4. Пилотная эксплуатация

- Интеграция PoC в боевую среду на ограниченный сегмент пользователей.
- Мониторинг дрейфа данных и комплаенс-контроль (152-ФЗ, 57580-1).
- Уточнение бюджета масштабирования.

5) Этап 5. Масштабирование

- Развёртывание промышленной инфраструктуры; внедрение CI/CD и MLOps конвейеров [Selectel-RD-2021].
- Резервирование ресурсов и план катастрофоустойчивости.

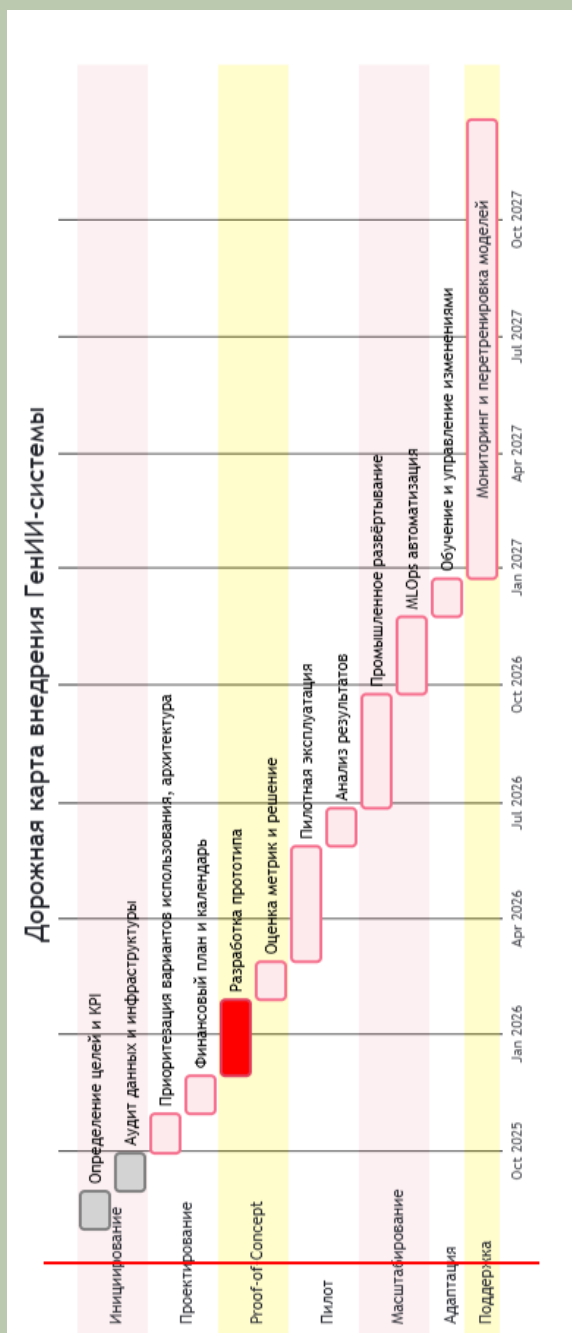
- Ключевое решение — финальная точка безубыточности (см. § 3.2).

6) Этап 6. Адаптация и изменения

- Программа обучения сотрудников; обновление регламентов.
- Коммуникационный план для минимизации сопротивления изменениям.
- Метрика: $\geq 70\%$ пользователей активно применяют ГенИИ-функции через 6 мес.

7) Этап 7. Поддержка и непрерывное улучшение

- Круглосуточный мониторинг модели, данных и инфраструктуры.
- Переобучение моделей по событию (дрейф $> 5\%$) либо по расписанию.
- Периодический пересмотр KPI и обновление дорожной карты.



4.2 Критерии выбора поставщика и технологии

Технические критерии

1) Сертификации и лицензии

- ФСТЭК/ФСБ (ИБ-работы, К1-данные).
- ISO 9001, ISO 27001, ГОСТ Р 57580.1 (финсектор).

2) Технологический стек

- Поддержка требуемых фреймворков (PyTorch 2.x, CUDA 12).
- Готовность к RAG-паттернам, встроенный векторный поиск.

3) Масштабируемость и отказоустойчивость

- Наличие GPU-кластеров $\geq$ A100/H100 с SLA 99,95%.

4) Интеграционные возможности

- API-совместимость (REST/gRPC), SDK, IaC-шаблоны.

5) Безопасность

- Поддержка KMS/HSM, сквозное шифрование данных «в покое» и «в транзите».

Коммерческие критерии

- ТСО-модель (CAPEX + OPEX) на горизонте $\geq$ 36 мес.
- Гибкость ценовой политики: почасовая тарификация GPU, фиксированные пакеты, SaaS-подписка.

- Финансовая устойчивость поставщика: коэффициент текущей ликвидности $\geq 1,5$; отсутствие убытков три последних года.
- Условия оплаты и штрафы: post-pay 30–45 дней, пеня за SLA-нарушения.
- Прозрачность коммерческих предложений (unit-экономика, расчёт TCO).

Репутационные критерии

- Портфолио и релевантные кейсы с публичными референсами.
- Отраслевые компетенции (oil&gas, финансы, госсектор).
- Наличие судебных споров и регуляторных штрафов — проверка открытых реестров.
- Социальная и экологическая ответственность (ISO 26000, ESG-отчёты).
- Кадровая стабильность: текучесть ключевых инженерных позиций $< 15\%$ в год.

Дорожная карта выбора

- Сбор требований → 2 нед.
- Длинный лист (RFI) → 3 нед.
- Короткий лист + презентации → 2 нед.
- Техничко-экономическое ранжирование → 1 нед.
- Демо-пилот финалистов → 4 нед.

- Финальные переговоры и контракт → 2 нед.
- Общий цикл: ~14 недель.

4.3 Управление рисками проекта

Идентификация рисков ГенИИ-проектов

Идентификация рисков представляет собой систематический процесс обнаружения, распознавания и документирования потенциальных угроз и возможностей, которые могут повлиять на достижение целей проекта внедрения генеративного искусственного интеллекта.

Основные категории рисков ГенИИ-проектов:

1) Технические риски

- Недостаточная производительность моделей для решения бизнес-задач
- Дрейф данных и деградация качества модели со временем
- Проблемы интеграции с существующими ИТ-системами
- Недоступность или нестабильность GPU-ресурсов

2) Операционные риски

- Превышение бюджета на вычислительные ресурсы
- Нехватка квалифицированных специалистов по машинному обучению
- Сопротивление персонала внедрению новых технологий

- Нарушение SLA по доступности системы

3) Регуляторные и комплаенс-риски

- Нарушение требований 152-ФЗ при обработке персональных данных
- Несоответствие отраслевым стандартам (ГОСТ Р 57580.1 для финсектора)
- Этические проблемы использования ИИ
- Изменения в законодательстве об искусственном интеллекте

4) Бизнес-риски

- Неопределённость ROI и сроков окупаемости
- Зависимость от внешних поставщиков облачных услуг
- Конкурентные риски при задержке внедрения
- Репутационные риски при сбоях ИИ-системы

Методы идентификации рисков:

- Мозговой штурм и (или) диверсионный анализ с участием междисциплинарной команды (ИТ, бизнес, юристы, специалисты по ИБ)
- Анализ исторических данных предыдущих ИТ-проектов и инцидентов
- SWOT-анализ для выявления внутренних и внешних факторов риска
- Экспертные интервью с ключевыми заинтересованными сторонами
- Чек-листы на основе лучших практик отрасли

Оценка и приоритезация рисков

Оценка рисков включает качественный и количественный анализ выявленных угроз с определением их приоритетности для разработки адекватных мер реагирования.

Качественная оценка рисков. Каждый риск оценивается по двум ключевым параметрам:

1) Вероятность возникновения (1 — 5 баллов):

- 1 — Очень низкая (менее 5 %)
- 2 — Низкая (5 — 25 %)
- 3 — Средняя (25 — 50 %)
- 4 — Высокая (50 — 75 %)
- 5 — Очень высокая (более 75 %)

2) Воздействие на проект (1 — 5 баллов):

- 1 — Очень низкое (незначительные задержки/затраты)
- 2 — Низкое (задержки до 1 месяца, затраты до 5 %)
- 3 — Среднее (задержки 1 — 3 месяца, затраты 5 — 15 %)
- 4 — Высокое (задержки 3 — 6 месяцев, затраты 15 — 30 %)
- 5 — Очень высокое (задержки более 6 месяцев, затраты свыше 30 %)

Количественная оценка рисков. Для критически важных рисков применяется количественная методология с расчётом потенциальных потерь:

- $SLE \text{ (Single Loss Expectancy)} = \text{Стоимость актива} \times \text{Фактор воздействия}$
- $ARO \text{ (Annual Rate of Occurrence)} = \text{Частота возникновения риска в год}$
- $ALE \text{ (Annual Loss Expectancy)} = SLE \times ARO$

Оценка и приоритезация рисков

Карта рисков представляет собой визуальное изображение рисков в координатах «вероятность — воздействие», позволяющее определить приоритеты для управления рисками.

Воздействие	Вероятность возникновения				
	Очень низкая (1)	Низкая (2)	Средняя (3)	Высокая (4)	Очень высокая (5)
Очень высокое (5)	5 	10 	15 	20 	25 
Высокое (4)	4 	8 	12 	16 	20 
Среднее (3)	3 	6 	9 	12 	15 
Низкое (2)	2 	4 	6 	8 	10 
Очень низкое (1)	1 	2 	3 	4 	5 

Интерпретация уровней риска:

 Низкий риск (1 — 5): Принимается без дополнительных мер

 Средний риск (6 — 15): Требуется мониторинг и планирование мер

 Высокий риск (16 — 25): Требуется немедленных действий по митигации

Стратегии управления основными рисками

Согласно методологии PMBoK, существует четыре основные стратегии реагирования на негативные риски:

- 1) Уклонение (Avoidance). Полное исключение риска путём изменения плана проекта.

Пример: Отказ от использования экспериментальных моделей ИИ в пользу проверенных решений для критически важных процессов.

- 2) Передача (Transfer). Перенос ответственности за риск на третью сторону.

Пример: Страхование киберрисков, использование облачных SLA с гарантиями доступности, аутсорсинг разработки специализированному подрядчику.

- 3) Снижение (Mitigation). Уменьшение вероятности возникновения или воздействия риска.

Пример: Создание резервных копий моделей, внедрение системы мониторинга дрейфа данных, обучение дополнительных специалистов.

- 4) Принятие (Acceptance). Сознательное принятие риска с созданием резервов.

Пример: Создание резерва бюджета 15 — 20 % на непредвиденные расходы по вычислительным ресурсам.

Планы управления основными рисками ГенИИ-проектов

Риск	Вероятность	Воздействие	Уровень	Стратегия	Меры митигации
Превышение бюджета на GPU-ресурсы	4	4	16 	Снижение	Гибридная модель развертывания, мониторинг затрат, автоматизированное
Дрейф данных и деградация модели	3	4	12 	Снижение	Система мониторинга качества, автоматическое переобучение, A/B-тестирование
Нехватка ML-специалистов	4	3	12 	Передача	Партнерство с вузами, аутсорсинг, использование AutoML-платформ
Нарушение требований 152-ФЗ	2	5	10 	Уклонение	Аттестованные облачные площадки, псевдонимизация данных, аудит ИБ

Риск	Вероятность	Воздействие	Уровень	Стратегия	Меры митигации
Сопrotивление персонала	3	3	 9	Снижение	Программа обучения, change-менеджмент, демонстрация пилотных успехов
Технические сбои интеграции	3	3	 9	Снижение	Позапная интеграция, тестовые среды, откат к предыдущей версии
Изменения в законодательстве об ИП	2	4	 8	Принятие	Мониторинг регуляторных изменений, гибкая архитектура, правовые консультации
Конкурентные риски	2	3	 6	Принятие	Ускорение time-to-market, фокус на уникальные преимущества

Мониторинг и контроль рисков

Управление рисками ГенИИ-проектов требует постоянного мониторинга и адаптации стратегий:

1) Ключевые индикаторы риска (KRI):

- Отклонение фактических затрат от плановых более чем на 15 %
- Снижение точности модели более чем на 5 % от базового уровня
- Увеличение времени отклика системы более чем на 50 %
- Текучесть ключевых специалистов более 20 % в квартал

2) Периодичность пересмотра:

- Еженедельные встречи по мониторингу рисков для критических проектов
- Ежемесячное обновление карты рисков
- Квартальный пересмотр стратегий управления рисками
- Годовой аудит эффективности риск-менеджмента

Эффективное управление рисками ГенИИ-проектов требует проактивного подхода, регулярного мониторинга и готовности к быстрой адаптации стратегий в условиях высокой технологической неопределённости.

4.4 Требования к техническому заданию

Общие положения

Техническое задание (ТЗ) — основной нормативный документ, определяющий требования к разработке программного обеспечения или автоматизированной системы.

Обязательные разделы технического задания

1) По ГОСТ 19.201—78 (для ПО):

- Введение
 - Цель и основания разработки
 - Область применения
- Назначение разработки
 - Функциональное назначение
 - Эксплуатационное назначение
 - Ожидаемые результаты
- Характеристика объектов автоматизации
 - Описание аппаратных и программных средств
 - Взаимодействие с другими системами
- Требования к программе (системе)
 - Функциональные требования
 - Нефункциональные требования (надежность, производительность, безопасность, совместимость)
 - Требования к интерфейсам
- Состав и содержание работ
 - Этапы и сроки разработки
- Порядок контроля и приемки
 - Методы испытаний
 - Критерии приемки
- Требования к подготовке объекта автоматизации к вводу в эксплуатацию

- Требования к документации
 - Источники разработки
- 2) По ГОСТ 34.602—2020 (для автоматизированных систем):
- Общие сведения
 - Назначение и цели создания системы
 - Характеристика объектов автоматизации
 - Требования к системе (функциональные, технические, информационные, эргономические, по безопасности)
 - Состав и содержание работ по созданию системы
 - Порядок контроля и приёмки системы
 - Требования к подготовке объекта автоматизации к вводу системы в действие
 - Требования к документированию
 - Источники разработки

Формулировки требований

- Требования должны быть конкретными, измеримыми, достижимыми, релевантными и ограниченными по времени (SMART).
- Функциональные требования: что система должна делать (например, поддержка определённых бизнес-процессов).
- Нефункциональные требования: как система должна работать (например, время отклика ≤ 1 сек, уровень доступности ≥ 99.95 %).

- Требования к интерфейсам, интеграции, совместимости с другими системами.
- Требования к безопасности, защите данных, соответствию нормативным актам (например, 152-ФЗ, ГОСТ Р 57580.1).

Критерии приёмки

- Соответствие всем функциональным и нефункциональным требованиям, изложенным в ТЗ.
- Успешное прохождение приёмочных испытаний (методика испытаний согласуется в ТЗ).
- Наличие и корректность всей сопроводительной документации.
- Отсутствие критических дефектов по результатам тестирования.
- Документальное подтверждение выполнения всех этапов работ (протоколы, акты приёмки).

Рекомендации по оформлению

- ТЗ оформляется на листах формата А4, с нумерацией страниц в верхней части листа.
- Лист утверждения и титульный лист — по ГОСТ 19.104—78.
- Внесение изменений — через дополнения к ТЗ.
- Для автоматизированных систем допускается оформление по ГОСТ 34.602—89 с учётом специфики отрасли.

4.5 Показатели эффективности и методы их измерения

Классификация KPI для ГенИИ-проектов

Показатели эффективности генеративных систем искусственного интеллекта должны охватывать технические, экономические и бизнес-аспекты внедрения. Система KPI строится по принципу сбалансированной системы показателей (BSC) с учётом специфики ИИ-технологий.

Основные группы показателей:

- Технические KPI — измеряют качество работы ИИ-системы
- Экономические KPI — оценивают финансовую эффективность
- Бизнес-KPI — отражают влияние на бизнес-процессы
- Пользовательские KPI — характеризуют принятие системы пользователями

Технические показатели эффективности

1) Качество модели:

- Точность (Accuracy) — доля правильных предсказаний от общего числа предсказаний
- Полнота (Recall) — доля найденных релевантных объектов от общего числа релевантных
- Прецизионность (Precision) — доля релевантных объектов среди найденных

- F1-мера — гармоническое среднее прецизионности и полноты

2) Производительность системы:

- Время отклика (Latency) — время от запроса до получения ответа
 - Целевое значение: ≤ 300 мс для интерактивных приложений
 - Измерение: медианное время отклика за период
- Пропускная способность (Throughput) — количество обработанных запросов в единицу времени
 - Целевое значение: ≥ 100 запросов/сек для корпоративных систем
- Доступность системы (Availability) — процент времени работоспособности
 - Целевое значение: $\geq 99,9\%$ (не более 8,76 часов простоя в год)

3) Стабильность и надёжность:

- Дрейф модели (Model Drift) — изменение качества модели со временем
 - Измерение: сравнение метрик качества с базовой линией
 - Пороговое значение: снижение точности не более чем на 5%
- Частота переобучения — количество обновлений модели в период
- Время восстановления после сбоя (MTTR) — среднее время устранения инцидентов

Экономические показатели эффективности

1) Показатели рентабельности:

- ROI (Return on Investment) — возврат инвестиций
 - Формула: $ROI = (Выгода - Затраты) / Затраты \times 100\%$
 - Целевое значение: $\geq 300\%$ за 3 года
- NPV (Net Present Value) — чистая приведённая стоимость
 - Целевое значение: $NPV > 0$ при ставке дисконтирования 15%
- Срок окупаемости (Payback Period)
 - Целевое значение: ≤ 18 месяцев для корпоративных проектов

2) Показатели затрат:

- TCO (Total Cost of Ownership) — совокупная стоимость владения
 - Измерение: CAPEX + OPEX за период жизненного цикла
- Стоимость обработки одного запроса
 - Формула: $Общие\ затраты / Количество\ обработанных\ запросов$
- Отклонение фактических затрат от плановых
 - Допустимое отклонение: $\pm 15\%$ от планового бюджета

Экономические показатели эффективности

1) Операционная эффективность:

- Автоматизация процессов — доля автоматизированных операций
 - Целевое значение: $\geq 70\%$ рутинных операций
- Сокращение времени выполнения процессов
 - Измерение: сравнение времени до и после внедрения ИИ
 - Целевое значение: сокращение на 40-60%
- Повышение качества продукции/услуг
 - Измерение: снижение количества ошибок, рекламаций
 - Целевое значение: снижение дефектов на 30-50%

2) Клиентские показатели:

- Удовлетворённость клиентов (CSAT)
 - Измерение: опросы клиентов по 5-балльной шкале
 - Целевое значение: $\geq 4,2$ балла
- Net Promoter Score (NPS)
 - Целевое значение: ≥ 50 для B2B, ≥ 30 для B2C
- Время решения обращений клиентов
 - Целевое значение: сокращение на 50-70%

Экономические показатели эффективности

1) Принятие системы:

- Коэффициент принятия (Adoption Rate):
 - Формула: $\frac{\text{Активные пользователи}}{\text{Общее количество пользователей}} \times 100\%$

- Целевое значение: $\geq 80\%$ через 6 месяцев после внедрения
 - Частота использования
 - Измерение: среднее количество сессий на пользователя в день
 - Время освоения системы (Time to Competency)
 - Целевое значение: ≤ 2 недель для базовых функций
- 2) Удовлетворённость пользователей:
- System Usability Scale (SUS)
 - Целевое значение: ≥ 70 баллов (хорошая юзабилити)
 - Количество обращений в техподдержку
 - Целевое значение: $\leq 5\%$ пользователей в месяц

Методики измерения показателей

1) Технические метрики:

- Автоматизированный мониторинг
 - Системы APM (Application Performance Monitoring)
 - Логирование всех запросов и ответов системы
 - Алерты при превышении пороговых значений
- A/B тестирование
 - Сравнение производительности разных версий модели
 - Статистическая значимость результатов ($p\text{-value} < 0,05$)
 - Размер выборки для достоверности результатов

2) Экономические метрики:

- Финансовый учёт
 - Детализация всех затрат по категориям (CAPEX/OPEX)
 - Ежемесячная сверка фактических и плановых расходов
 - Учёт скрытых затрат (обучение, интеграция)
- Бизнес-кейс трекинг
 - Измерение фактических выгод vs прогнозируемых
 - Корректировка ROI-расчётов на основе реальных данных
 - Анализ чувствительности к изменению ключевых параметров

3) Бизнес-метрики:

- Процессный анализ
 - Хронометраж бизнес-процессов до и после внедрения
 - Анализ качества выходных данных процессов
 - Измерение производительности сотрудников
- Клиентская аналитика
 - Регулярные опросы удовлетворённости
 - Анализ поведенческих данных (веб-аналитика, CRM)
 - Когортный анализ изменения клиентских метрик

Дашборды и отчётность

- Операционный дашборд (ежедневный мониторинг):
 - Доступность системы в реальном времени
 - Количество обработанных запросов за день
 - Среднее время отклика
 - Количество ошибок и инцидентов
- Тактический дашборд (еженедельная отчётность):
 - Тренды ключевых технических метрик
 - Анализ пользовательской активности
 - Отклонения от плановых показателей
 - Статус выполнения проектных задач
- Стратегический дашборд (ежемесячная отчётность):
 - ROI и финансовые показатели
 - Достижение бизнес-целей проекта
 - Прогресс по дорожной карте
 - Рекомендации по оптимизации

Периодичность измерений и пересмотра KPI

Группа показателей	Частота измерения	Частота пересмотра целевых значений
Технические	В реальном времени / ежедневно	Ежемесячно
Экономические	Еженедельно / ежемесячно	Ежеквартально
Бизнес-показатели	Еженедельно / ежемесячно	Ежеквартально
Пользовательские	Еженедельно	Ежемесячно

Критерии успешности проекта:

- Достижение 80 % от запланированных KPI через 12 месяцев
- Положительный ROI через 18 месяцев
- Принятие системы ≥ 70 % целевых пользователей
- Отсутствие критических инцидентов более 4 часов

Эффективная система измерения показателей должна обеспечивать баланс между детальностью мониторинга и практичностью использования, позволяя принимать своевременные управленческие решения для оптимизации ГенИИ-системы.

4.6 Методика расчёта экономической эффективности

Область применения методики

Важное примечание: Оценка экономической эффективности в первую очередь необходима негосударственным корпоративным заказчикам, осуществляющим инвестиции в системы ГенИИ с целью получения прибыли и повышения конкурентоспособности.

Для государственных заказчиков — органов государственного управления — оценка экономической эффективности может оказаться не критичной, поскольку их основной целью является повышение качества предоставляемых государственных услуг, обеспечение национальной безопасности и выполнение социальных функций. В таких случаях основное внимание уделяется соответствию бюджетным ограничениям и достижению поставленных социально-экономических целей.

Особенности применения для различных типов заказчиков

1) Корпоративные заказчики:

- Обязательный расчёт ROI, NPV, IRR
- Анализ срока окупаемости инвестиций
- Сравнение альтернативных вариантов инвестиций
- Учёт налоговых последствий и амортизации

2) Государственные заказчики:

- Фокус на достижении целевых показателей госпрограмм
- Соответствие бюджетным ограничениям
- Оценка социально-экономического эффекта
- Анализ бюджетной эффективности

Концептуальные основы экономической эффективности ГенИИ-проектов

Оценка экономической эффективности генеративных систем искусственного интеллекта представляет собой комплексный процесс, включающий анализ финансовых показателей, учитывающих временную стоимость денег и риски, связанные с внедрением передовых технологий. В отличие от традиционных ИТ-проектов, ГенИИ-системы характеризуются высокой неопределённостью результатов, длительными периодами внедрения и существенными первоначальными инвестициями.

Методология экономической оценки ГенИИ-проектов базируется на трёх ключевых показателях: возврате инвестиций (ROI), чистой приведённой стоимости (NPV) и внутренней норме доходности (IRR). Каждый из этих показателей представляет собой отдельный аспект для анализа финансовой привлекательности проекта и помогает принимать обоснованные инвестиционные решения.

Особенностью ГенИИ-проектов является их способность генерировать как прямую, так и косвенную экономическую выгоду. Прямая выгода включает снижение операционных затрат, увеличение производительности и повышение качества продукции. Косвенная выгода может проявляться в улучшении клиентского опыта, ускорении инновационных процессов и создании новых источников дохода.

Формула расчёта возврата инвестиций (ROI)

Возврат инвестиций является базовым показателем эффективности, демонстрирующим отношение полученной прибыли к вложенным средствам. Для ГенИИ-проектов формула ROI адаптируется с учётом специфики технологии:

$$\text{ROI} = (\text{Общая выгода от ГенИИ} - \text{Общие затраты на ГенИИ}) / \text{Общие затраты на ГенИИ} \times 100 \%$$

где:

- Общая выгода от ГенИИ включает прямую экономию затрат, увеличение доходов, повышение

производительности и стоимость рисков, которых удалось избежать

- Общие затраты на ГЕНИИ охватывают капитальные расходы, операционные затраты, обучение персонала и скрытые издержки

Исследования показывают, что средний ROI ГЕНИИ-проектов составляет 3.5 рубля на каждый вложенный рубль, при этом 5 % ведущих организаций достигают ROI в 8 рублей на рубль инвестиций. Критическим фактором успеха является правильное определение временных рамок расчёта: слишком короткий период может не отражать долгосрочные выгоды, а слишком длинный увеличивает неопределённость прогнозов.

Методология расчёта чистой приведённой стоимости (NPV)

Чистая приведённая стоимость представляет собой разность между приведённой стоимостью денежных притоков и оттоков, дисконтированных по определённой ставке. Для ГЕНИИ-проектов формула NPV имеет следующий вид:

$$NPV = \sum (CF(t) / (1 + r)^t) - I_0$$

где:

- $CF(t)$ — чистый денежный поток в период t
- r — ставка дисконтирования
- t — временной период
- I_0 — первоначальные инвестиции

Выбор ставки дисконтирования для ГенИИ-проектов требует особого внимания. В российских условиях рекомендуется использовать ставку 15 — 20 %, что отражает повышенные риски инновационных технологий по сравнению с традиционными ИТ-проектами. Для государственных проектов может применяться ставка 10 — 15 %, учитывающая социальную значимость и долгосрочную перспективу.

Положительное значение NPV указывает на экономическую целесообразность проекта, при этом проекты с большим значением NPV являются более привлекательными. Критическим моментом является корректное прогнозирование денежных потоков, которое должно учитывать постепенное наращивание эффекта от использования ГенИИ-системы.

Расчёт внутренней нормы доходности (IRR)

Внутренняя норма доходности представляет собой ставку дисконтирования, при которой NPV проекта равна нулю. IRR показывает максимально допустимый уровень затрат на привлечение капитала для финансирования проекта.

IRR определяется из уравнения:
$$\sum (CF(t) / (1 + IRR)^t) = I_0$$

Для практических расчётов IRR используется итерационный метод или встроенные функции электронных таблиц. Проект считается привлекательным, если IRR превышает требуемую норму доходности (cost of capital).

В российских условиях пороговые значения IRR для ГенИИИ-проектов составляют:

- Стартапы: 40 — 50 %
- Корпоративные проекты: 25 — 30 %
- Государственные проекты: 15 — 20 %

Важно отметить, что IRR может иметь множественные значения для проектов с нестандартными денежными потоками, что требует дополнительного анализа с использованием модифицированной внутренней нормы доходности (MIRR).

Индекс прибыльности (PI)

Индекс прибыльности дополняет анализ эффективности, показывая отношение приведённой стоимости денежных притоков к размеру инвестиций:

$$PI = \sum(CF(t) / (1 + r)^t) / I_0$$

Проект считается эффективным при $PI > 1$, что эквивалентно положительному значению NPV. Индекс прибыльности особенно полезен при ранжировании альтернативных проектов с ограниченным бюджетом.

Пошаговые расчёты для типовых сценариев

Сценарий 1: Автоматизация службы поддержки средней компании

Исходные данные:

- Первоначальные инвестиции: 5 000 000 рублей

- Экономия на зарплате операторов: 2 400 000 рублей в год
- Операционные расходы: 600 000 рублей в год
- Период проекта: 5 лет
- Ставка дисконтирования: 15 %

Расчёт чистых денежных потоков:

- Годовой чистый денежный поток = 2 400 000 – 600 000 = 1 800 000 рублей

Расчёт NPV:

- $NPV = -5\,000\,000 + 1\,800\,000/(1.15)^1 + 1\,800\,000/(1.15)^2 + 1\,800\,000/(1.15)^3 + 1\,800\,000/(1.15)^4 + 1\,800\,000/(1.15)^5$
- $NPV = -5\,000\,000 + 1\,565\,217 + 1\,361\,058 + 1\,183\,529 + 1\,029\,591 + 895\,297$
- $NPV = 1,034,692$ рубля

Расчёт ROI:

- Общая выгода за 5 лет = 1 800 000 × 5 = 9 000 000 рублей
- $ROI = (9\,000\,000 - 5\,000\,000) / 5\,000\,000 \times 100 \% = 80 \%$

Расчёт IRR:

- Решая уравнение $NPV = 0$, получаем $IRR \approx 24.3 \%$

Вывод: Проект экономически эффективен ($NPV > 0$, $IRR >$ требуемая доходность 15 %).

Сценарий 2: Система генерации контента для маркетинга

Исходные данные:

- Первоначальные инвестиции: 15 000 000 рублей
- Увеличение продаж от улучшенного контента: 8 000 000 рублей в год
- Экономия на создании контента: 3 000 000 рублей в год
- Дополнительные операционные расходы: 2 000 000 рублей в год
- Период проекта: 4 года
- Ставка дисконтирования: 18 %

Расчёт чистых денежных потоков:

- Годовой чистый денежный поток = 8 000 000 + 3 000 000 – 2 000 000 = 9 000 000 рублей

Расчёт NPV:

- $NPV = -15\,000\,000 + 9\,000\,000/(1.18)^1 + 9\,000\,000/(1.18)^2 + 9\,000\,000/(1.18)^3 + 9\,000\,000/(1.18)^4$
- $NPV = -15\,000\,000 + 7\,627\,119 + 6\,463\,661 + 5\,478\,543 + 4\,642\,783$
- $NPV = 9\,212\,106$ рублей

Расчёт PI:

- $PI = (7\,627\,119 + 6\,463\,661 + 5\,478\,543 + 4\,642\,783) / 15\,000\,000 = 1.61$

Расчёт простого срока окупаемости:

- $\text{Срок окупаемости} = 15\,000\,000 / 9\,000\,000 = 1.67$ года

Вывод: Высокоэффективный проект с быстрой окупаемостью

Сценарий 3: Крупномасштабная система анализа данных предприятия

Исходные данные:

Первоначальные инвестиции: 80 000 000 рублей

- Денежные потоки по годам:
 - Год 1: 5 000 000 рублей (внедрение, низкая эффективность)
 - Год 2: 20 000 000 рублей (наращивание эффекта)
 - Год 3: 35 000 000 рублей (полная эффективность)
 - Год 4: 40 000 000 рублей (масштабирование)
 - Год 5: 45 000 000 рублей (оптимизация)
- Ставка дисконтирования: 16 %

Расчёт NPV:

- $NPV = -80\,000\,000 + 5\,000\,000/(1.16)^1 + 20\,000\,000/(1.16)^2 + 35\,000\,000/(1.16)^3 + 40\,000\,000/(1.16)^4 + 45\,000\,000/(1.16)^5$
- $NPV = -80\,000\,000 + 4\,310\,345 + 14\,881\,657 + 22\,399\,249 + 22\,322\,538 + 21\,426\,434$
- $NPV = 5\,340\,223$ рубля

Расчёт дисконтированного срока окупаемости:

- Накопленный NPV к концу 3-го года: $-80\,000\,000 + 4\,310\,345 + 14\,881\,657 + 22\,399\,249 = -38\,408\,749$
- Накопленный NPV к концу 4-го года: $-38\,408\,749 + 22\,322\,538 = -16\,086\,211$
- Окупаемость наступает в 5-м году

Факторы, влияющие на точность расчётов

1) Системные риски оценки

- При расчёте экономической эффективности ГенИИ-проектов необходимо учитывать специфические факторы неопределённости:
- Технологические риски — возможность появления более эффективных решений
- Регуляторные риски — изменения в законодательстве о данных и ИИ
- Компетенционные риски — недостаток квалифицированного персонала
- Интеграционные риски — сложности внедрения в существующую инфраструктуру

2) Рекомендуемые корректировки

- Для повышения точности оценок рекомендуется:
- Увеличивать ставку дисконтирования на 2 — 5 % для учёта технологических рисков
- Включать в расчёты резерв на непредвиденные расходы в размере 15 — 25 %
- Использовать консервативные прогнозы выгод в первые 2 года внедрения
- Проводить анализ чувствительности для ключевых параметров

Правильная оценка экономической эффективности ГенИИ-проектов требует комплексного подхода, сочетающего традиционные финансовые методы с учётом специфики инновационных технологий. Только такой подход позволяет принимать обоснованные

инвестиционные решения и максимизировать экономический эффект от внедрения генеративного искусственного интеллекта.

Список использованных источников

1. Вечер В. С., Никольский В. П., Стегайлов В. В. (2016) Соотношение между энергопотреблением и производительностью для GPU-алгоритмов молекулярной динамики // Сб. тр. конф. «Суперкомпьютерные дни в России». Москва, 2016.
2. [ITmatika-Platform] Выбор облачной платформы для бизнеса — Айтиматика. 2020. — URL: <https://clck.ru/3Mvetu> (дата обращения: 03.07.2025).
3. [TAdviser-GenAI] Генеративный искусственный интеллект [Электронный ресурс] // TAdviser. — 2025. — URL: <https://clck.ru/3MpfaF> (дата обращения: 26.06.2025).
4. [FSHybrid] Гибридный облачный ЦОД: плюсы & минусы и сценарии применения / FS Community. — URL: <https://clck.ru/3MtYqu> (дата обращения: 01.07.2025).
5. [ГОСТ 19.101] Единая система программной документации. Виды программ и программных документов.
6. [ГОСТ 19.201] Единая система программной документации. Техническое задание. Требования к содержанию и оформлению.
7. [ГОСТ 27.202] ГОСТ Р 27.202—2012: Надёжность в технике. Управление надёжностью. Стоимость жизненного цикла.

8. [ГОСТ 34.602] Техническое задание на создание автоматизированной системы.
9. [ГОСТ 12207] ГОСТ Р ИСО/МЭК 12207—2010: Системная и программная инженерия. Процессы жизненного цикла программных средств.
10. [ГОСТ 56923] ГОСТ Р 56923—2016: Системная и программная инженерия. Управление жизненным циклом. Часть 3. Руководство по применению ИСО/МЭК 12207.
11. [ГОСТ 57193] ГОСТ Р 57193—2016: Системная и программная инженерия. Процессы жизненного цикла систем.
12. [ГОСТ 58302] ГОСТ Р 58302—2018: Управление стоимостью жизненного цикла. Номенклатура показателей для оценивания стоимости жизненного цикла изделия. Общие требования.
13. [ГОСТ 58535] ГОСТ Р 58535—2019: Стоимостной инжиниринг. Термины и определения.
14. [ГОСТ 59793] ГОСТ Р 59793—2021: Автоматизированные системы. Стадии создания.
15. [ГОСТ 60300] ГОСТ Р МЭК 60300-3-3—2021: Надёжность в технике. Менеджмент надёжности. Стоимость жизненного цикла.
16. [AzureReliability] Документация по надёжности Azure. — URL: <https://clck.ru/3MtYoF> (дата обращения: 01.07.2025).
17. [VKCloud-152] Защита персональных данных в облаке: что нужно знать по 152-ФЗ — VK Cloud. 2019. — URL: <https://clck.ru/3MvewY> (дата обращения: 03.07.2025).

18. [YandexSec] Информационная безопасность: как защищать данные... / Yandex Cloud Blog. — URL: <https://clck.ru/3MtYrX> (дата обращения: 01.07.2025).
19. [IEMag-Hidden-Costs] ИТ-затраты явные и скрытые [Электронный ресурс] // Intelligent Enterprise. — URL: <https://clck.ru/3MqWVo> (дата обращения: 28.06.2025).
20. [DailyMoscow-2024] Как внедрить ИИ в малый и средний бизнес. Люди, деньги, инструменты [Электронный ресурс] // Daily Moscow. — 2024. — URL: <https://clck.ru/3Mpgiz> (дата обращения: 26.06.2025).
21. [ITGLOBAL-VMwareCloud] Как мигрировать между облаками на VMware // ITGLOBAL.COM, 28.05.2025. — URL: <https://clck.ru/3MwefC> (дата обращения: 04.07.2025).
22. [TU-Don-2023] Как не потерять деньги при внедрении WMS [Электронный ресурс] // Технологии учёта. — 2023. — URL: <https://clck.ru/3MqWmJ> (дата обращения: 28.06.2025).
23. [Computerra-2025] Ключевые ошибки при внедрении ИИ: как не загубить бизнес [Электронный ресурс] // Computerra. — 2025. — URL: <https://clck.ru/3MqWin> (дата обращения: 28.06.2025).
24. **Кокотов Д. В.** (2023) Методика организации микросервисной архитектуры // Информатика и вычислительная техника. 2023.
25. **Кукушко А.** (2024) Как создать дорожную карту для внедрения ИИ и добиться результата. LinkedIn Pulse, 23.02.2024. — URL: <https://clck.ru/3MxNqx> (дата обращения: 05.07.2025).

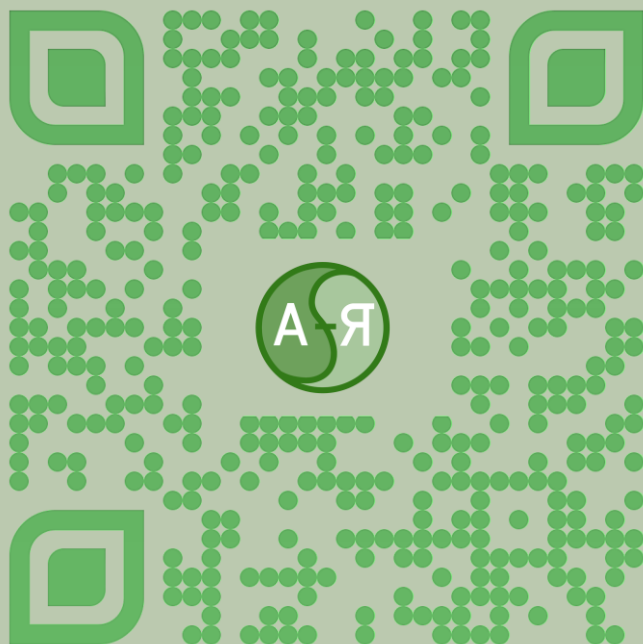
26. [CloudMTS-GOST] Облако CloudMTS подтвердило соответствие ГОСТ Р 57580.1-2017 — ComNews. 2023. — URL: <https://clck.ru/3Mvexh> (дата обращения: 03.07.2025).
27. [T1Cloud-GOST] Облако T1 Cloud соответствует требованиям безопасности для финансового сектора — T1 Cloud. 2023. — URL: <https://clck.ru/3Mvf2Y> (дата обращения: 03.07.2025).
28. [K2Cloud-GOST] Облачная платформа K2 Cloud соответствует требованиям ГОСТ Р 57580.1—2017 — CNews. 2025. — URL: <https://clck.ru/3Mveyrn> (дата обращения: 03.07.2025).
29. [MS-Azure-ML] Операции машинного обучения (MLOps v2). Azure Architecture Center. 2024. — URL: <https://clck.ru/3MxNrh> (дата обращения: 05.07.2025).
30. [TBank-OPEX] Операционные расходы: что это такое, какие затраты входят в структуру [Электронный ресурс] // Т-Банк. — 2025. — URL: <https://clck.ru/3Mprj6H> (дата обращения: 27.06.2025).
31. [SelectelPublic] Описание продукта «Публичное облако на базе VMware» / Selectel. — URL: <https://clck.ru/3MtYmz> (дата обращения: 01.07.2025).
32. **Попов П. А.** (2020) Основные аспекты нецелесообразности миграции КИС в облачную среду // Информ. системы, № 2, 2020.
33. [AI-Success-CNews] От чего зависит успех проекта в области ИИ — CNews Conferences. 2022. — URL: <https://clck.ru/3Mvf5H> (дата обращения: 03.07.2025).

34. **Степанова Г. А.** (2023) Постоянные и переменные затраты на внедрение и поддержку ERP-систем и их расчёт [Электронный ресурс] // Корпоративные информационные системы. — 2023. — № 21. — URL: <https://clck.ru/3Mpgts> (дата обращения: 26.06.2025).
35. [Listohod-SED] Проблемы внедрения СЭД: 6 ключевых ошибок и решения [Электронный ресурс] // Listohod. — 2025. — URL: <https://clck.ru/3MqXB4> (дата обращения: 28.06.2025).
36. [Reeng-2025] Прямые и косвенные затраты на эксплуатацию оборудования [Электронный ресурс] // Reeng. — 2025. — URL: <https://clck.ru/3MqUXv> (дата обращения: 28.06.2025).
37. [SiglaVision-BI] Скрытая стоимость BI: что не учитывают 8 из 10 компаний при внедрении систем бизнес-аналитики [Электронный ресурс] // SiglaVision. — URL: <https://clck.ru/3MqUcG> (дата обращения: 28.06.2025).
38. [Runet-Compliance] Служить и защищать: как соответствовать нормативным требованиям — Runet.News. 2025. — URL: <https://clck.ru/3Mvf3U> (дата обращения: 03.07.2025).
39. [Selectel-RD-2021] Собственная методология разработки R&D-проектов в AI. Блог Selectel, 09.09.2021. — URL: <https://clck.ru/3MxNsu> (дата обращения: 05.07.2025).
40. [CBM-Solutions] Стоимость переноса данных 1С в Москве [Электронный ресурс] // КУБ 1С. — 2025. — URL: <https://clck.ru/3MqX6f> (дата обращения: 28.06.2025).

41. [Komiinform-DataMig] Сценарии миграции с зарубежных решений на российские ПО // Комиинформ, 04.06.2025. — URL: <https://clck.ru/3Mweg9> (дата обращения: 04.07.2025).
42. [Games4Business] Тренинг «Управление изменениями» [Электронный ресурс] // Лаборатория Деловых Игр. — URL: <https://clck.ru/3MqX9a> (дата обращения: 28.06.2025).
43. [Upperator-2023] Цена искусственного интеллекта для логистических компаний [Электронный ресурс] // Upperator. — 2023. — URL: <https://clck.ru/3MpgCe> (дата обращения: 26.06.2025).
44. [ServerSpace-Checklist] Чек-лист: как выбрать облачного провайдера? — Serverspace.kz. 2024. — URL: <https://clck.ru/3Mverc> (дата обращения: 03.07.2025).
45. [TrendMicro] Что такое соответствие требованиям для облака? / Trend Micro. — URL: <https://clck.ru/3MtYpK> (дата обращения: 01.07.2025).
46. [ITGradPrivate] Что такое частное облако (Private Cloud) / ИТ-ГРАД. — URL: <https://clck.ru/3MtYpj> (дата обращения: 01.07.2025).
47. [FedAG-ROI] Экономический эффект от внедрения ИИ [Электронный ресурс] // FedAG. — 2025. — URL: <https://clck.ru/3MpjCc> (дата обращения: 27.06.2025).
48. [ComplianceCloud] «Compliance в облаках. Не бойтесь высоты...» / Compliance Control. — URL: <https://clck.ru/3MtYp2> (дата обращения: 01.07.2025).

49. [Cloud4Y-2019] GPU сервер для машинного обучения и работы с ИИ [Электронный ресурс] // Cloud4Y. — 2019. — URL: <https://clck.ru/3MpgLA> (дата обращения: 26.06.2025).
50. [ISO 15686-5:2017] ISO 15686-5: Здания и сооружения. Планирование срока службы. Часть 5. Оценка стоимости жизненного цикла.
51. [TCO-2023] Conceptual Framework for Urban Farming (UF) Cost Components: The Application of Iceberg Total Cost of Ownership (TCO) Model. International Journal of Business and Technology Management. Published online September 1, 2023. — DOI: 10.55057/ijbtm.2023.5.s2.13.
52. **Mazur B., Wasiak M.** (2018) TCO — Total Cost of Ownership as a supporting tool of choosing the car fleet in rental company. WUT Journal of Transportation Engineering. 2018; 123:95-109. — DOI: 10.5604/01.3001.0013.7465.
53. **Mieritz L., Kirwin B.** (2005) Defining Gartner Total Cost of Ownership. Gartner Group, 8 December 2005.
54. [Nubes-GOST] Nubes подтвердил соответствие облака требованиям ГОСТ Р 57580.1—2017. 2024. — URL: <https://clck.ru/3Mvezh> (дата обращения: 03.07.2025).
55. **Stroparo T. R., Floriani N.** (2024) Sustainable Development Goals (SDGS) and Total Cost of Ownership (TCO): a holistic approach. Revista De Gestão E Secretariado, 15(9), e4236. — DOI: 10.7769/gesec.v15i9.4236.

КОНТАКТЫ



Адрес: 129110, г. Москва, ул. Гиляровского, д. 48,
оф. 100



Телефон: +7 (909) 695-41-38



Email: info@aia.expert