

Проблематика при внедрении GenAI: Интеграция с базой знаний

Олег Зельдин

Алекс Берг

**В начале вспомним
про стандарты**

Управление контентом документации КЦ (СОРС)



В контактном центре необходимо документировать многие процессы, процедуры и регламенты. Важно обеспечить ряд ключевых характеристик процесса управления контентом для того, чтобы обеспечить корректность, единообразность и проверяемость выполнения процессов и действий как на уровне операторов, так и на уровне менеджеров КЦ

Управление контентом документации КЦ. Ключевые драйверы

- Авторизация персонала для изменения контента и утверждения изменений
- Сбор обратной связи от пользователей и действия по результатам
- Периодическая ревизия контента на предмет актуальности, точности и ясности
- Скрытие устаревшего контента. Сохранение версионности, истории изменений
- Порядок приоритизации изменений контента
- Управление консистентностью между каналами (включая каналы БПЧ)

Управление контентом должно применяться ко всем документам, используемым в КЦ. Особое внимание: база знаний, скрипты операторов, регламенты процесса контроля качества. Сам по себе процесс управления контентом должен быть документирован с описанием конкретики по каждому драйверу, указанному выше



**Разница между
традиционным и
GenAI в контексте
управления знаниями**

Важно понимать разницу между AI и GenAI в контексте управления знаниями.

AI

ИИ в управлении знаниями фокусируется на организации, извлечении и анализе данных для оптимизации процессов и повышения эффективности

Примеры решаемых задач:

- Маркировка контента
- Выявление закономерностей
- Ответы на базовые запросы

AI опирается на распознавание закономерностей и анализ данных, чтобы помочь вам быстро и точно принимать обоснованные решения

GenAI

Генерация нового контента. Он синтезирует информацию и генерирует оригинальные ответы, документы или идеи на основании правил и алгоритмов, полученных в результате анализа существующих данных

Примеры решаемых задач:

- Составление отчетов
- Создание резюме
- Предоставление индивидуальных ответов на сложные вопросы

GenAI добавляет уровень креативности и динамической генерации контента, преобразуя способ управления и использования организационных знаний.

Граф знаний (Knowledge Graph)

Хайп-цикл Гартнера

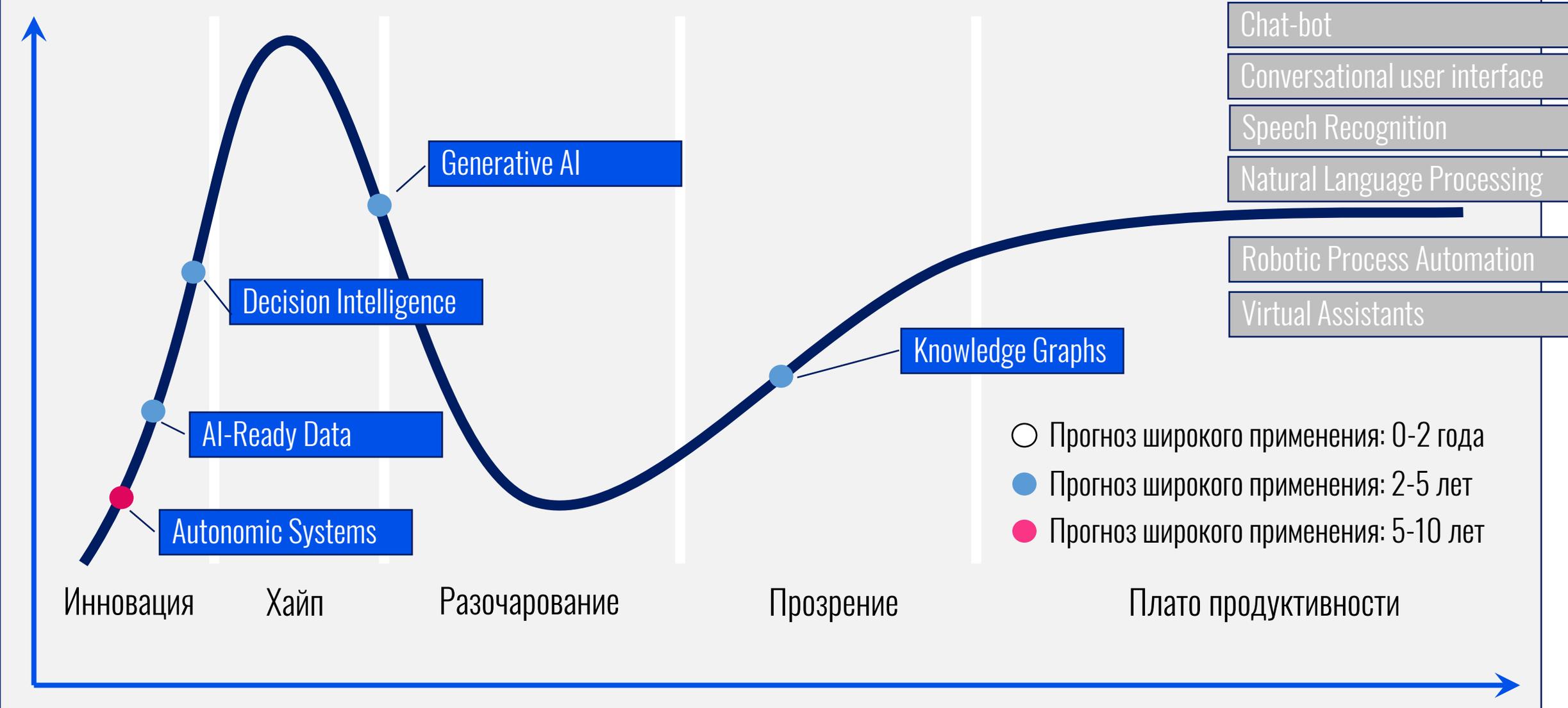
ИСТОЧНИК:



<https://clck.ru/3GbLgT>

Gartner[®]

Хайп-цикл Gartner. ИИ для КЦ. 2024

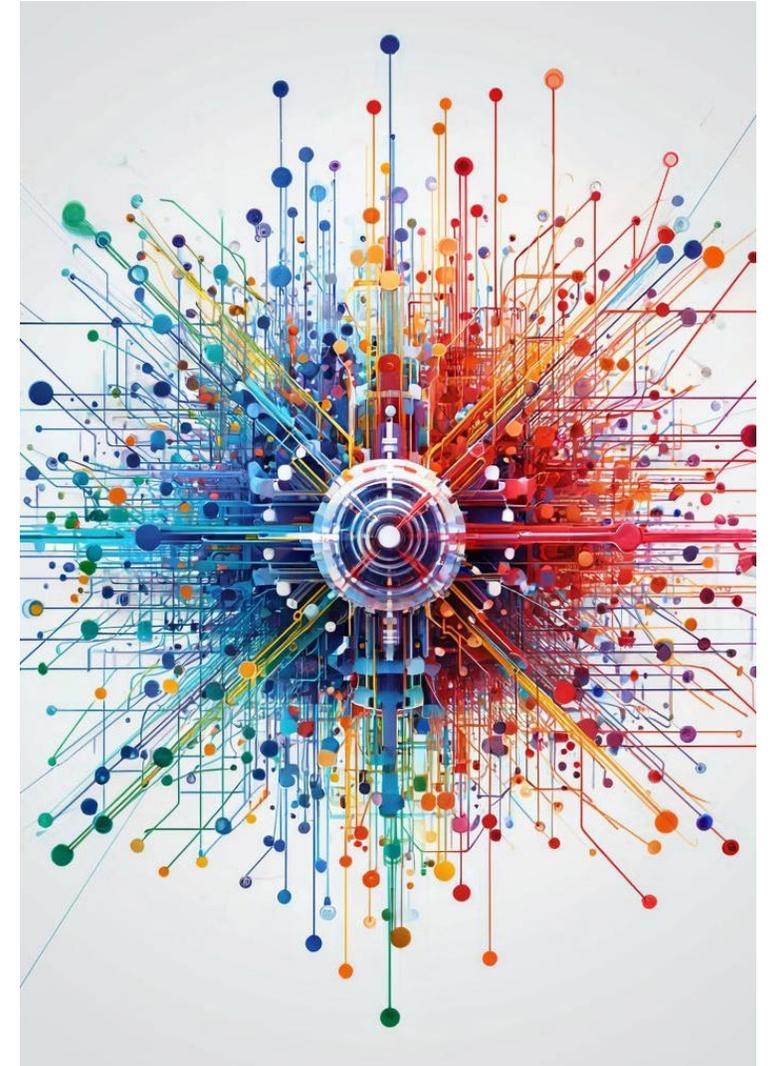


Выводы

1. На международном уровне многие технологии традиционного искусственного интеллекта, применяемые в КЦ (такие как чат-боты, интерфейсы для взаимодействия с пользователем, распознавание речи, обработка естественного языка, автоматизация процессов с помощью роботов и виртуальные ассистенты), уже не являются новыми и внедряются повсеместно. КЦ, в которых эти технологии не используются – уже в числе отстающих.
2. В настоящее время новыми считаются следующие технологии:
 - **Генеративный ИИ** и продвинутые текстовые и голосовые боты, созданные на его основе.
 - **Графы знаний** – в частности, их применение к базам знаний для операторов и клиентским данным (выявление рисков, мошенничества и т.д.).
 - **Интеллект решений** – следующий этап после предиктивной аналитики, который помогает принимать основанные на фактах и просчитанные решения в сложных и изменяющихся условиях.
 - **Данные, готовые для ИИ** – качественные и структурированные данные с учетом контекста, значительно повышающие эффективность ИИ. Например, работа с базами знаний.
 - **Автономные системы** – системы, которые полностью функционируют без человеческого вмешательства.

Что такое граф знаний (Knowledge Graph)?

- **Граф знаний (Knowledge Graph)** — это способ организации информации в виде **связей между сущностями**, который позволяет ИИ понимать не просто тексты, а **смыслы и отношения между вещами**.
- Граф знаний — это как **умная карта** с «точками» (объектами) и «линиями» (связями между ними). Эти связи говорят ИИ: кто есть кто, что с чем связано и как.



👉 Пример (в тексте):

«Александр Пушкин родился в Москве. Он был поэтом и автором “Евгения Онегина”.»

📌 Граф знаний превращает это в:

- Александр Пушкин → **родился в** → Москва
- Александр Пушкин → **является** → поэт
- Александр Пушкин → **написал** → “Евгений Онегин»

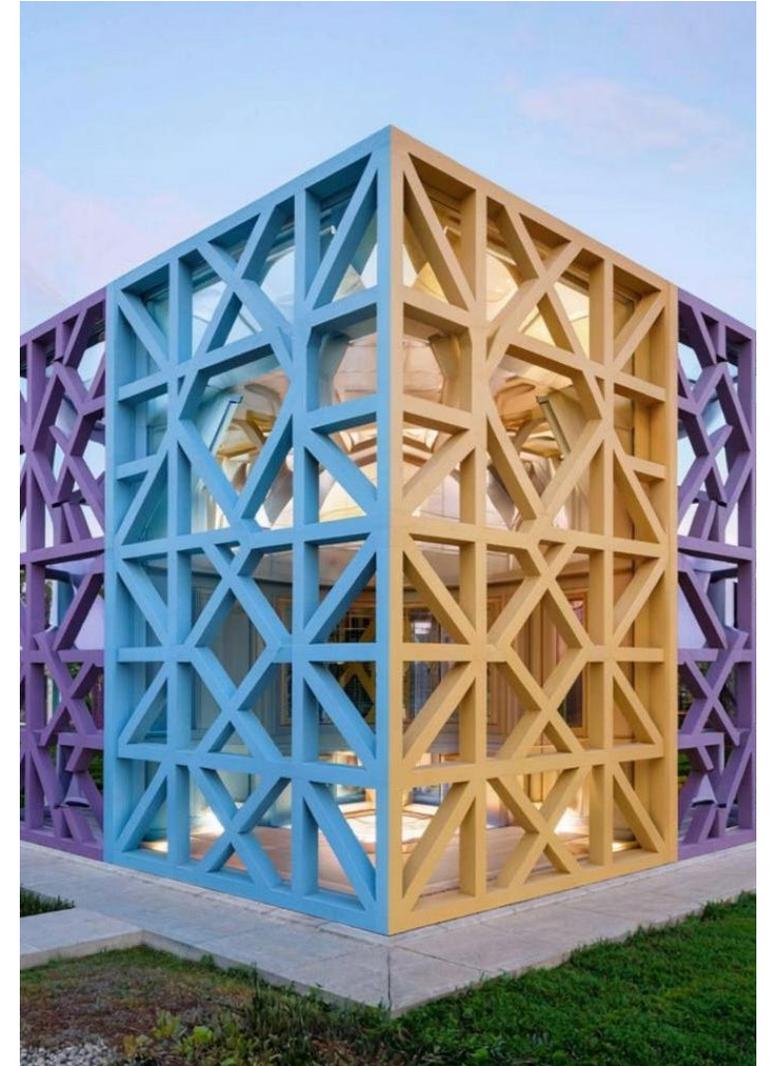
Это и есть граф — узлы (сущности) и **связи** между ними.





Состав графа знаний

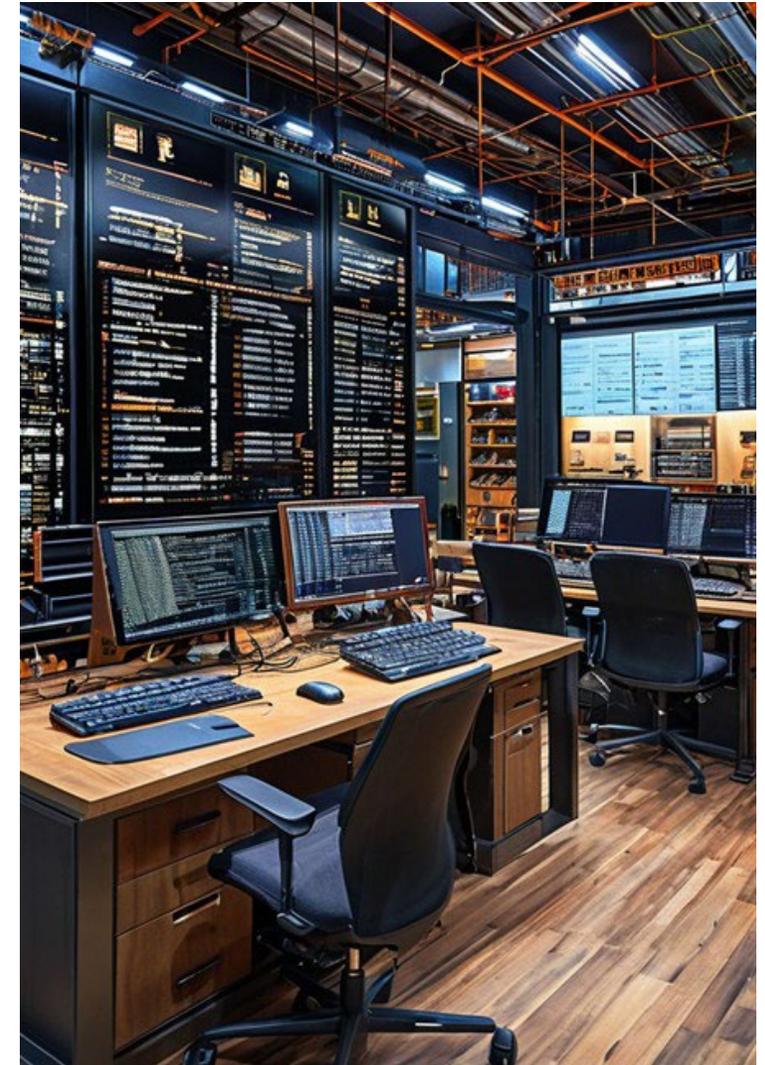
Компонент	Что это
Узел (node)	Объект или сущность (человек, продукт, дата, событие)
Связь (edge)	Отношение между двумя узлами (например: “работает в”, “написал”, “часть”)
Атрибуты	Дополнительная информация (например: дата рождения, категория, тег)





Где используется граф знаний?

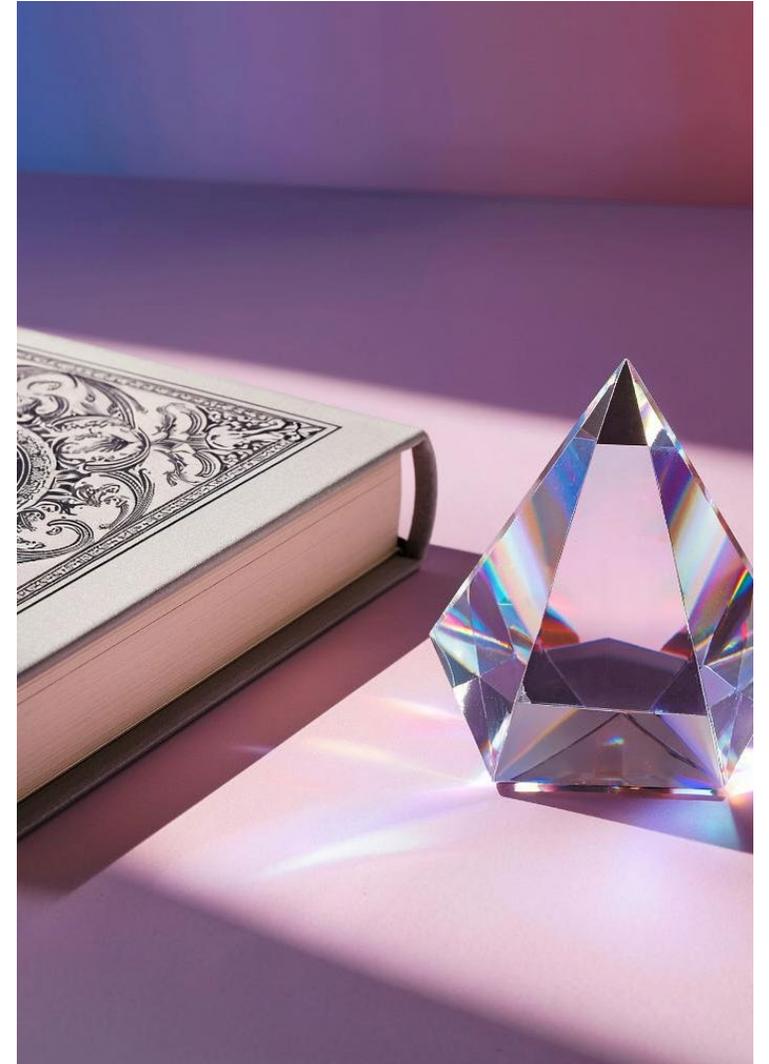
Область	Как помогает
Поисковые системы	Понимание смысла запроса. Пример: Google показывает карточку с фактами.
Чаты и GenAI	Помогает связать факты, не запутаться в похожих именах, уточнить контекст.
Рекомендательные системы	Например: кто смотрел этот фильм — интересовался этим актёром.
Контактные центры	Умный поиск ответов: "Если клиент спрашивает про доставку в Москву — ищи правила доставки по городам + сроки."





Отличие графа от базы знаний

База знаний	Граф знаний
Это просто документы или статьи	Это сеть фактов и связей между ними
Нужен поиск по ключевым словам	Возможен смысловой поиск и вывод связей
Подходит для FAQ	Подходит для сложных запросов и анализа





Пример для контактного центра

Запрос: "Клиент хочет вернуть товар по акции"

С помощью графа знаний:

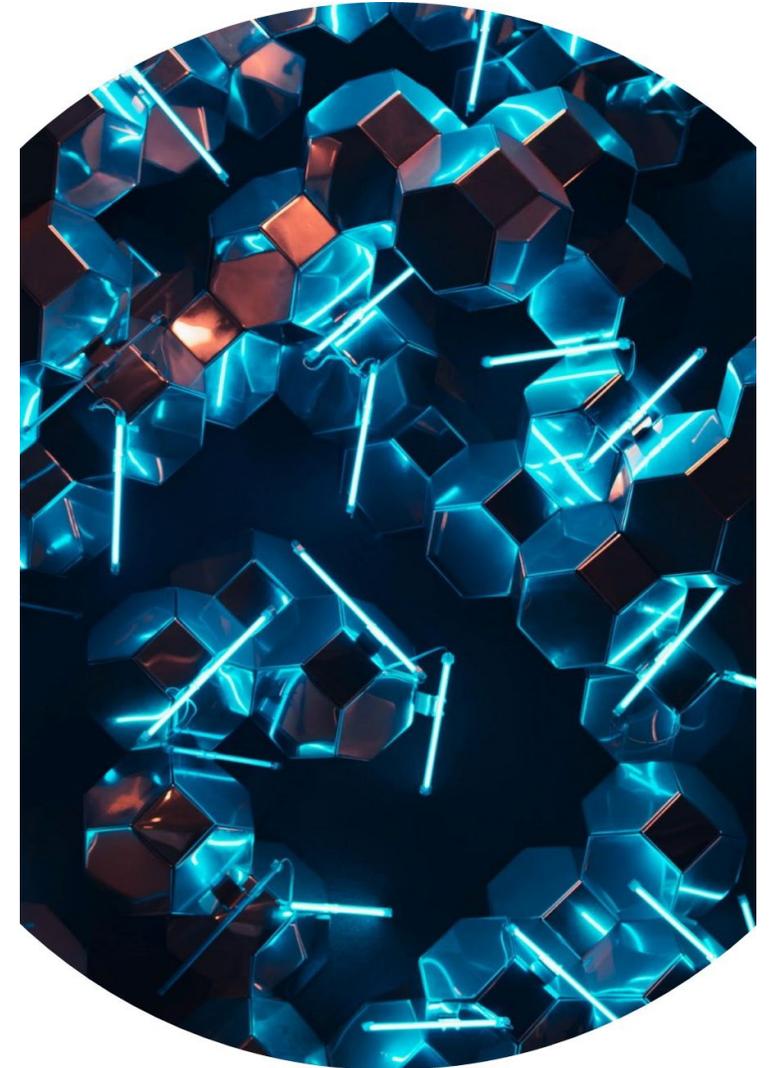
- Клиент → **совершил покупку** → Товар А
- Товар А → **участвовал в** → Акция В
- Акция В → **имеет условие** → Товар не подлежит возврату

✓ GenAI скажет:

“По условиям акции В, возврат невозможен.”

GenAI может:

- использовать графы знаний для **уточнения фактов**
- автоматически **строить графы из текстов**
- **объяснять связи** между понятиями



Структурированные и неструктурированные данные

Что такое структурированные данные?

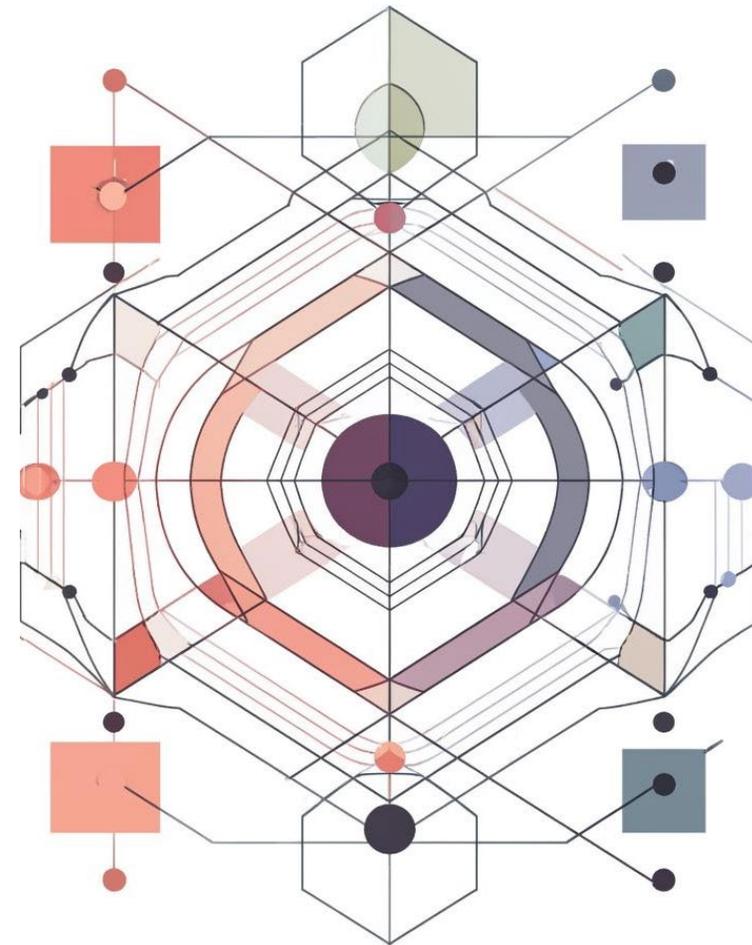
Структурированные данные — это информация, которая **организована по чёткой схеме**, понятной как человеку, так и компьютеру. У таких данных всегда есть **фиксированный формат**: строки, столбцы, поля, теги и т. д.

 **Пример – таблица Excel:**

Имя клиента	Номер телефона	Дата обращения	Причина
Иванов И.И.	8-900-123-45-67	01.06.2025	Вопрос по оплате
Сидорова А.А.	8-901-777-88-99	02.06.2025	Техническая ошибка

→ Это **структурированные данные**, потому что:

- Каждая строка — это отдельная запись (например, обращение клиента)
- Каждая колонка — это чётко определённое поле



Важно!

- Структурированные данные легко анализировать, фильтровать, сортировать
- Они хорошо подходят для поиска, обучения моделей ИИ, отчётов и визуализации

📁 Примеры форматов структурированных данных:

- Таблицы в Excel или Google Sheets
- Базы данных (MySQL, PostgreSQL и др.)
- JSON, XML, CSV с чёткой структурой
- Отметки и поля в CRM-системах (например, имя, дата обращения, тег обращения)

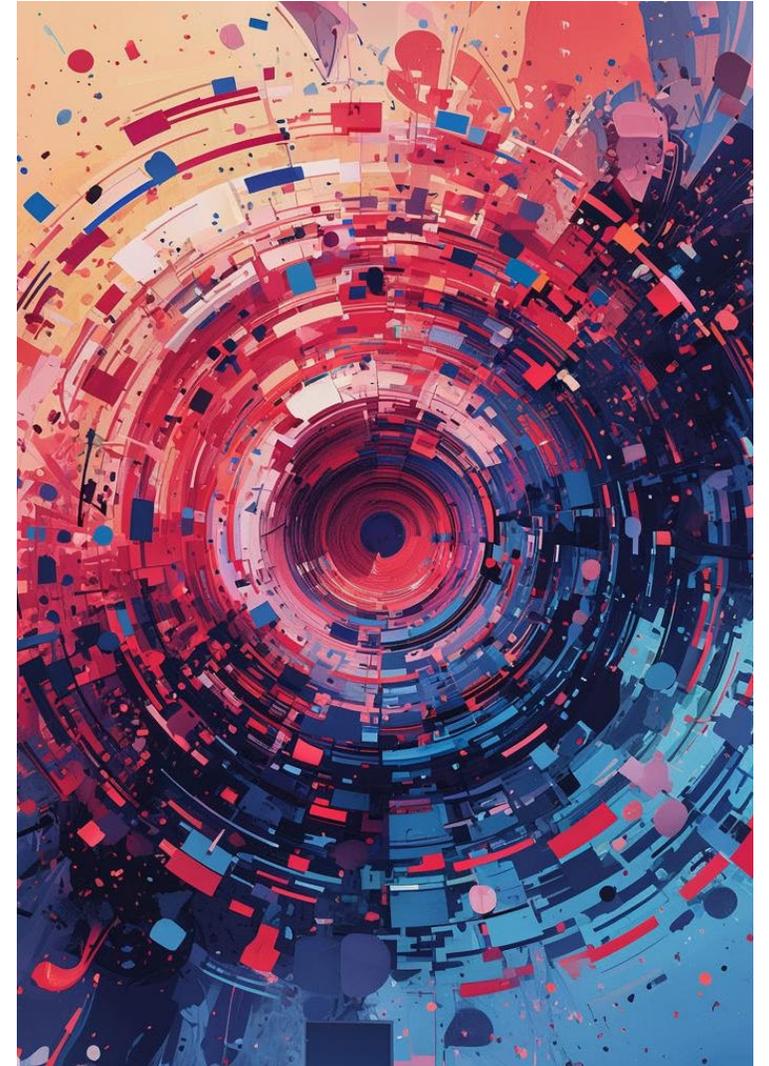


✘ Антипод — неструктурированные данные:

📌 Пример:

- "Клиент Иванов позвонил в пятницу. Жаловался на проблемы с оплатой. Мы пообещали разобраться.«

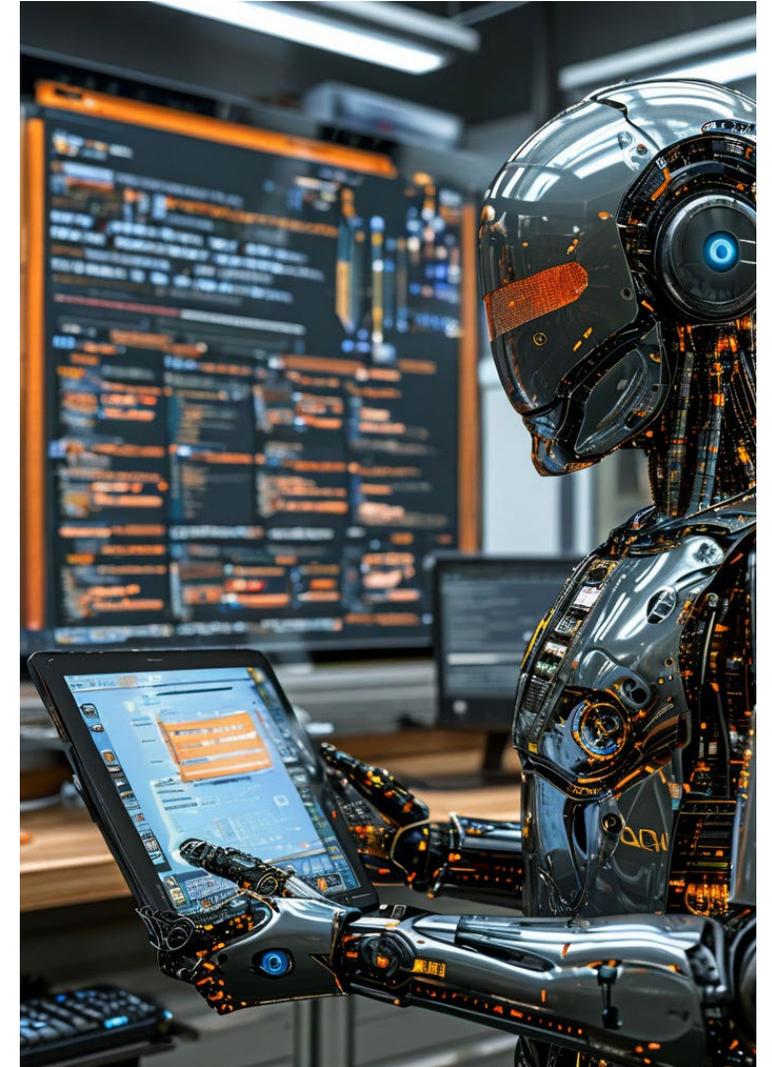
Это неструктурированный текст — его нельзя просто так поместить в таблицу без предварительной обработки (например, извлечения имени, темы и даты).





GenAI и структурированные данные

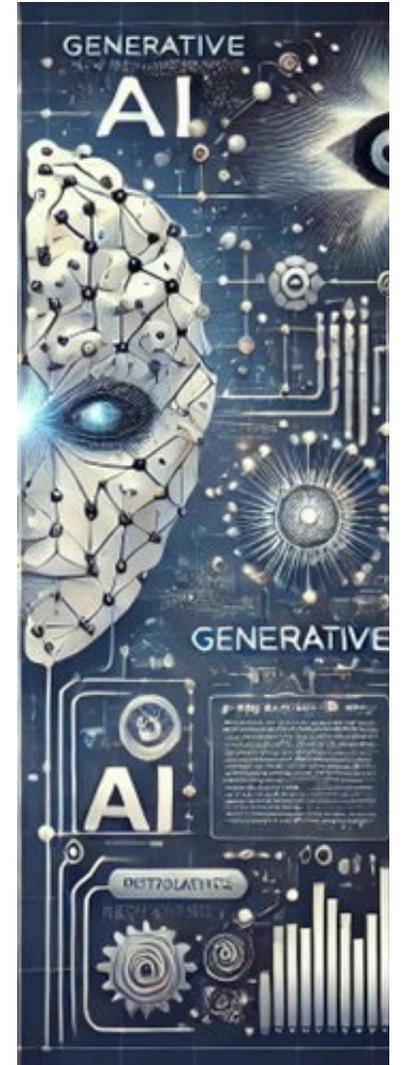
- Хотя GenAI умеет работать с неструктурированными данными, обучать его гораздо легче на структурированных
- GenAI лучше работает, если у него есть доступ к структурированной базе знаний (например, FAQ с полями: вопрос, ответ, теги).
- Такие данные легче встраивать в архитектуру RAG (поиск + генерация).



Архитектура RAG. (Retrieval-Augmented Generation)

Что такое RAG?

- **RAG (Retrieval - Augmented Generation)** — это архитектурный подход в работе генеративного ИИ, который **объединяет генерацию текста с извлечением внешних данных**
- GenAI (например, ChatGPT) обычно отвечает на основе того, чему он научен. Но если ему нужно ответить на вопрос, связанный с актуальной, специфической или внутренней информацией (например, политика компании, база знаний КЦ) — он может "не знать". **RAG решает эту проблему:** сначала ищется релевантная информация в базе знаний, а затем модель использует её для генерации точного ответа.





Как работает RAG?

Retrieval (извлечение):

- Сначала специальный компонент ищет наиболее подходящие документы из базы знаний или внутреннего хранилища (напр., FAQ, инструкции, статьи). Обычно используется векторный поиск (по смыслу, а не по ключевым словам).

Augmentation (обогащение)

- Найденные фрагменты добавляются в prompt (входной запрос) модели.

Generation (генерация)

- Модель (например, GPT-4) получает запрос + найденные тексты и генерирует ответ, опираясь на реальные данные, а не только на «память».

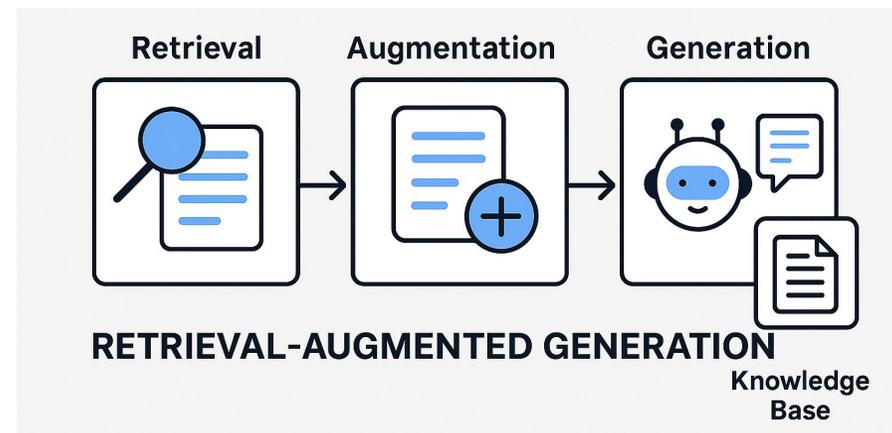
Пример

Без RAG:

- *"Как перевести звонок в другую очередь?"*
- Ответ: *"Извините, я не знаю."*

С RAG:

- **Retrieval**: нашёл статью: *"Чтобы перевести звонок, нажмите #9, выберите нужную очередь, затем Enter."*
- **Generation**: *"Чтобы перевести звонок, нажмите #9, выберите нужную очередь, затем подтвердите с помощью Enter."*



✅ Преимущества и ⚠️ Подводные камни RAG

- ✓ Подключение к актуальным и внутренним данным
 - ✓ Меньше «галлюцинаций»
 - ✓ Возможность внедрения в корпоративную среду без дообучения модели
 - ✓ Простая замена контента (обновил статьи — обновились ответы)
-
- ✓ Нужна чистая и структурированная база знаний
 - ✓ Важно обеспечить качественный поиск — иначе будут находиться нерелевантные данные
 - ✓ Требуется обработка длинных документов, разбиение на фрагменты, векторизация и т.д.



Технологические аспекты внедрения GenAI

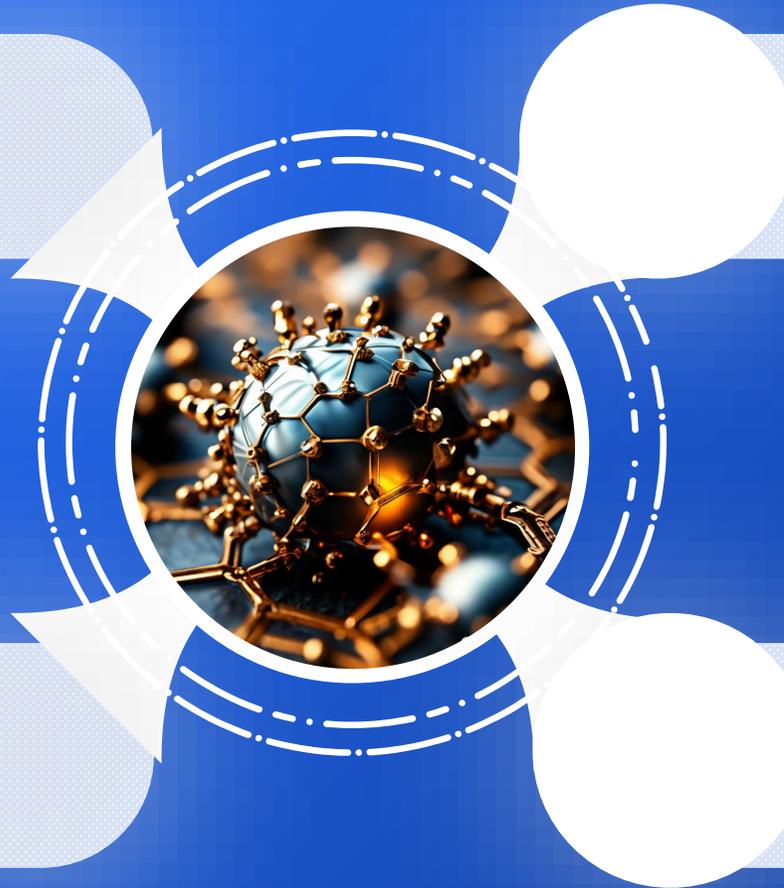
Технологические аспекты внедрения GenAI

Базовая технология

Модель внедрения

Интеграция с каналами и
данными

Технические проблемы
и ограничения



Интеграция с каналами и данными

Интеграция с корпоративными источниками включает базы знаний, скрипты и CRM-системы.

Используется **подход GenAI + NLP + Prompt Engineering** для добавления контекстной информации к запросам.

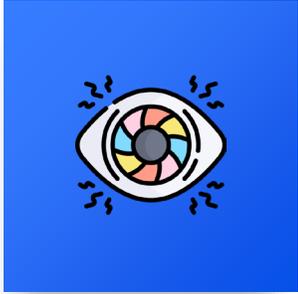
Например, перед запросом система может извлекать статьи из базы знаний для повышения точности ответов.

Интеграция с голосовыми каналами для распознавания речи клиента, передачи её модели и генерации ответа, который затем синтезируется в речь в реальном времени.

Комбинации (ASR + LLM + TTS) начали внедряться в 2023 году в голосовых ботах.



Технические проблемы и ограничения



Галлюцинации ГИИ

- GenAI может сгенерировать правдоподобно звучащий, но неверный ответ
- Необходимость технологических решений против галлюцинаций: фильтры и валидация ответов, ограничения рамками известных фактов



Интеграция в устаревшие системы

- Без достаточной **цифровой зрелости проектов** GenAI легко увязнуть в попытках «подружить» новые технологии со старыми платформами
- Необходимо готовить системы соответствующим образом (CRM, БЗ, Телефония)



Скорость и масштабируемость

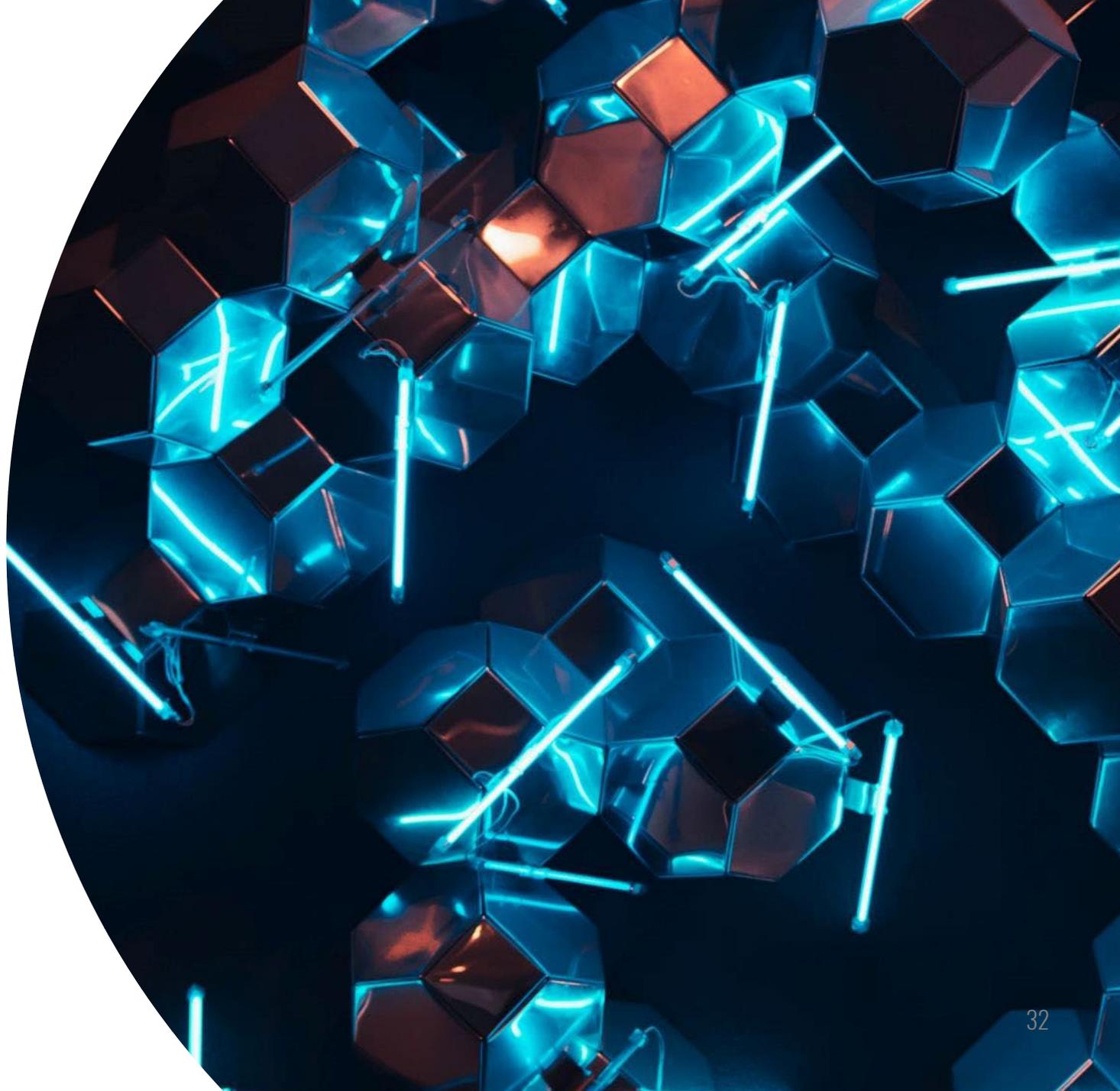
- Крупные модели требуют значительных вычислительных ресурсов. Обработка тысячи одновременных чатов через LLM – нетривиальная задача с точки зрения масштабирования
- Используется ограничение нагрузки, смешанная архитектура

Технологии ГИИ – от демо к промышленным пилотам

В целом, технологическая база ГИИ для КЦ к концу 2024 г. вышла из стадии опытных демо в стадию промышленных пилотов. Компании теперь накапливают компетенции, как наиболее эффективно применять эти инструменты. Следующий шаг – выработка стандартов качества, надежности и безопасности GenAI-решений, чтобы они могли работать стабильно в масштабах десятков тысяч обращений в день.

Необходимо инвестировать в подготовку данных и знаний

- Если внутренние базы данных устарели или разрознены, то ИИ не сможет получать точную информацию, что приведет к ошибкам. Очевидно, что проекты GenAI следует начинать с основ данных: объединения информационных силосов и обновления знаний.
- Качество ответов GenAI прямо пропорционально качеству знаний, на которых он обучен. Поэтому ранний урок – **привести в порядок базы знаний**, FAQ, скрипты, оцифровать ранее неформализованную экспертизу

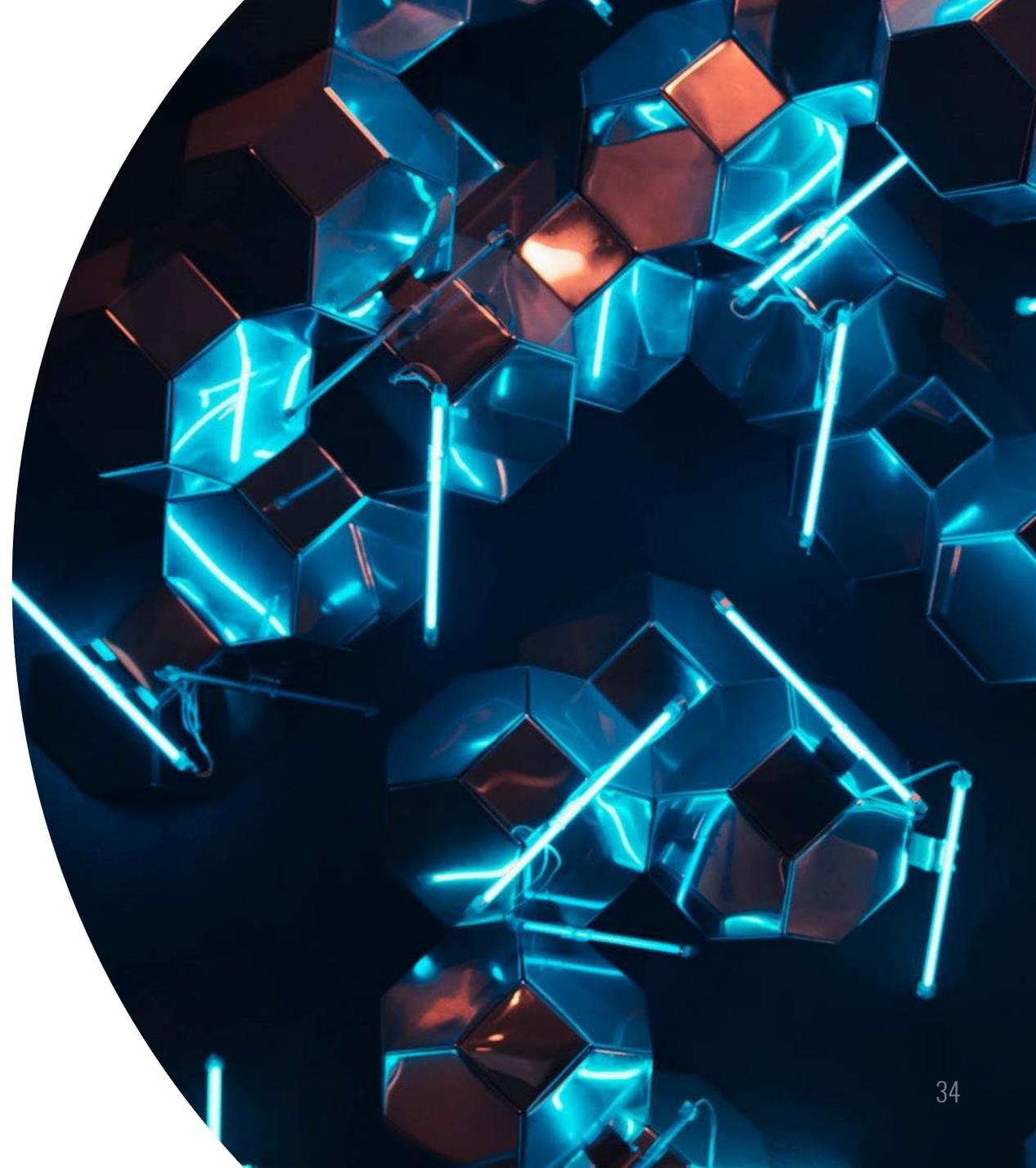


Внедрение GenAI в КЦ. Уроки первых внедрений от вендора базы знаний

GenAI. Проблемы первых внедрений

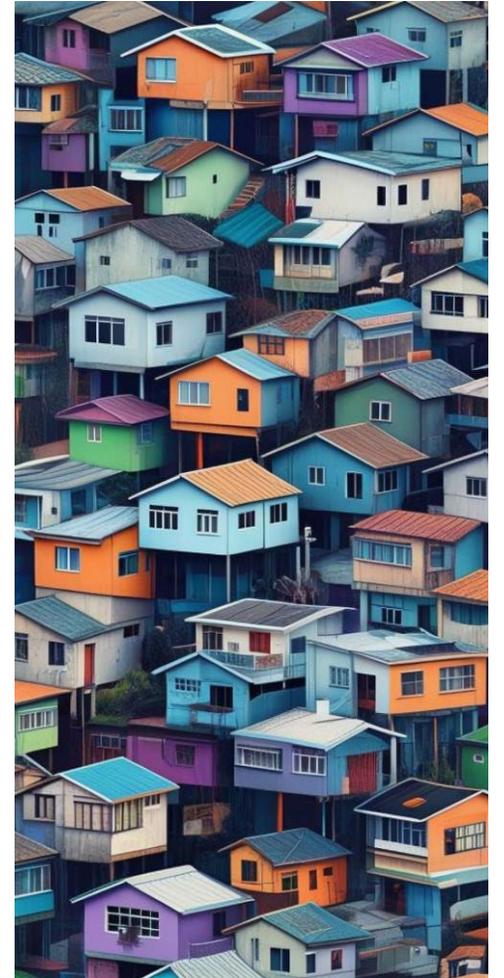
- Большинство первых проектов по внедрению GenAI в сервисе не достигли заявленных целей ни по снижению операционных расходов, ни по улучшению качества клиентского опыта.

- Исследование Gartner указывает, что в 2025 году 100% проектов виртуальных помощников (virtual assistant) в сфере обслуживания клиентов, не интегрированных с современными системами управления знаниями (Knowledge Management), не смогут справиться со своими задачами!



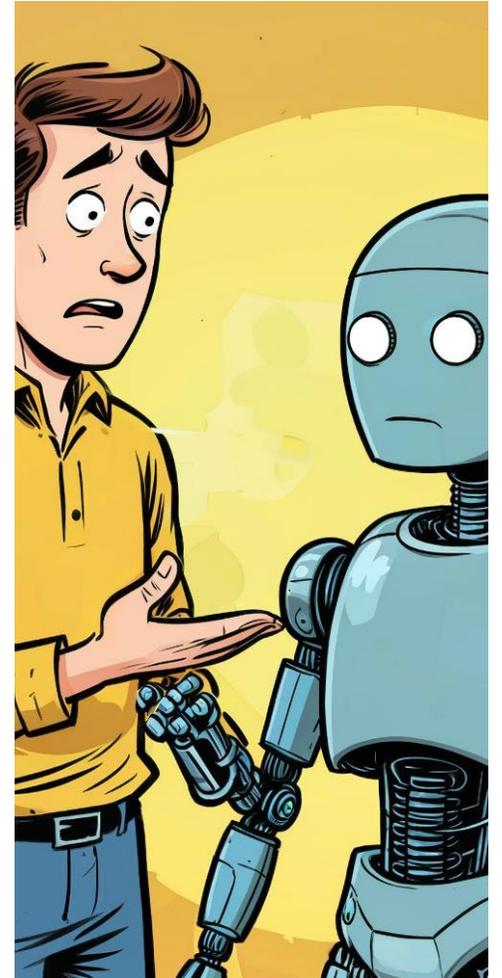
Существование нескольких изолированных хранилищ знаний и контента, на которых обучают GenAI

- Эффективность GenAI напрямую зависит от **качества и релевантности** содержимого, которое он получает.
- В обслуживании клиентов все ответы должны быть конкретно привязаны к специфике бизнеса, к его продуктам и предложениям.
- Если знания и данные разбросаны по нескольким «хранилищам», GenAI может получать противоречивые вводные, без единой системы проверки и подтверждения корректности.
- В результате выводы (ответы) ИИ становятся **недостаточно надёжными**.



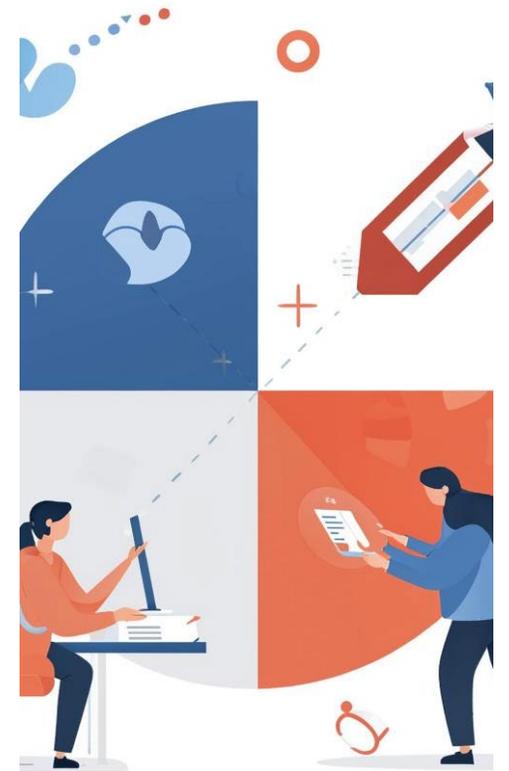
Отсутствие продуманной системы управления «промптом» (prompt management)

- Чтобы Генеративный ИИ выдавал действительно ценные результаты, необходимо грамотно управлять промптом — то есть формировать такие запросы к ИИ, которые соответствуют бизнес-целям.
- По сути, это похоже на роль «супервайзера» для нового сотрудника, дающего чёткие рабочие инструкции.
- В современных системах управления знаниями должна присутствовать библиотека лучших практик по созданию промптов, а также удобные средства быстрой настройки под текущие потребности бизнеса.



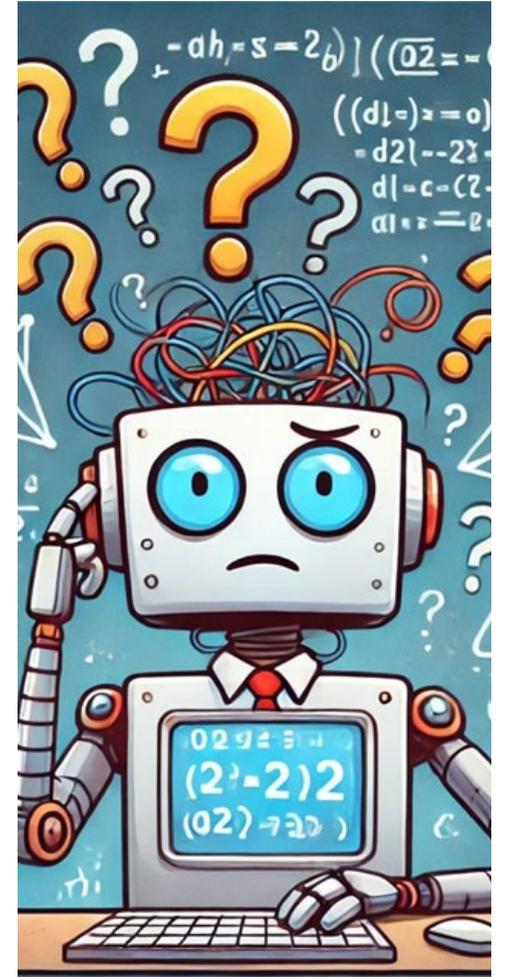
Непроработанные механизмы соблюдения регламентов и контроля качества пользовательского опыта

- Любая GenAI-система должна работать в рамках установленных, проверяемых правил.
- Например, необходимо **исключать** из анализа данные, связанные со строгим нормативным регулированием, если возникает риск несоблюдения комплаенса.
- Без детальных настроек (fine-grained controls) невозможно предоставлять массовые «достоверные» ответы в разных каналах и для разных сегментов клиентов.
- Более того, когда подобные правила срабатывают в процессе диалога, часто отсутствует **плавный «шаг назад»**: пользователь не понимает, что происходит, а компании не удаётся корректно подхватить разговор. Клиенты раздражаются, а негатив может быстро выйти в соцсети.



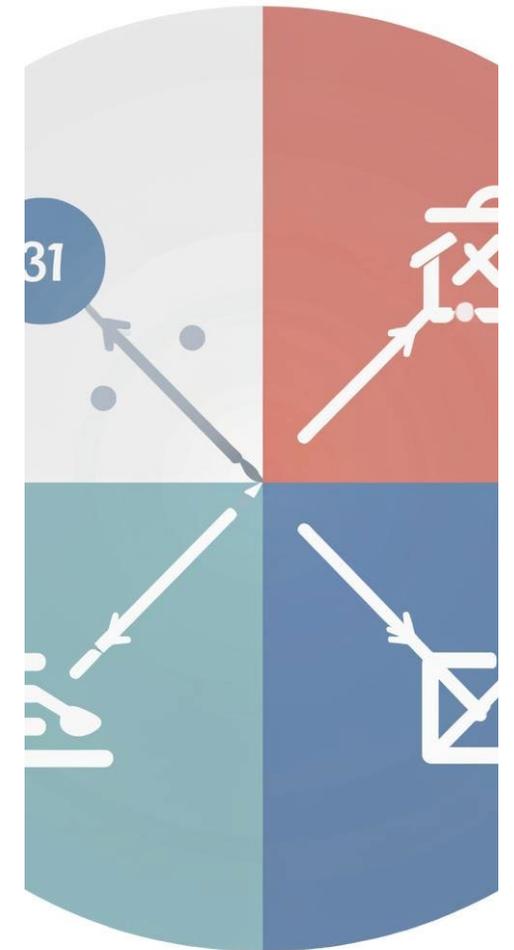
Отсутствие надёжных механизмов контроля качества выходных данных GenAI

- Генеративный ИИ склонен к так называемой «галлюцинации» (hallucination) — когда модель генерирует ошибочные или нерелевантные ответы.
- Если не организовать надёжную систему проверки качества в реальном времени, обеспечение точности и уместности ответов становится проблемой.
- Пара-тройка неправильных ответов в процессе общения с клиентом уже слишком много, особенно с учётом того, что люди сейчас весьма настороженно относятся к ответам «от ИИ». На ранних этапах часто недооценивали важность этой задачи, что приводило к негативному клиентскому опыту



Недостаток сквозной аналитики и «замкнутого цикла» (Closed-Loop Analytics)

- Чтобы непрерывно улучшать работу Генеративного ИИ, нужны механизмы **отслеживания** и **оценки** взаимодействий с ним
- Необходимо понимать, насколько эффективны использованные промпты, анализировать реакции пользователей (в том числе неявные сигналы), чтобы своевременно вносить корректировки.
- Без этого GenAI может надолго застрять на уровне «очередного примитивного чат-бота», который не оправдал ожиданий.



Рекомендации для успешного внедрения GenAI

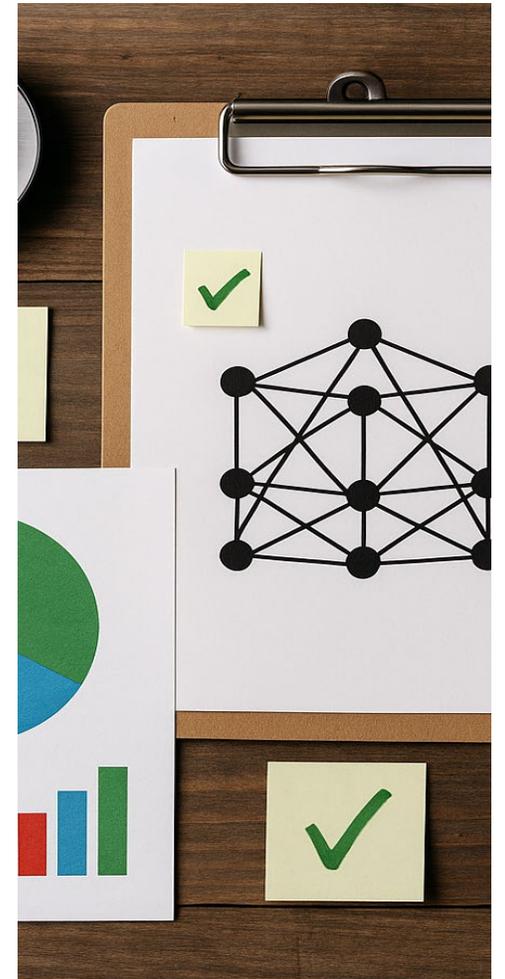
Аудит содержания

 **Что это значит:** Проверить, что база знаний не устарела, в ней нет ошибок, лишнего или дублирующего контента.

 **Пример:** В базе знаний есть инструкция по возврату товара, но она написана 3 года назад и не соответствует текущим правилам. Если GenAI ссылается на неё — он выдаст неверный ответ.

Что делать:

- Удалить или обновить старые документы
- Проверить, нет ли противоречий
- Убедиться, что информация покрывает все типовые вопросы клиентов



Обогатите неструктурированные данные для создания качественного контекста для GenAI

Разработка контролируемых словарей с глоссариями:

- Создайте стандартизированный список терминов (глоссарий), чтобы обеспечить единообразие в понимании специфических для вашей области терминов внутри организации.

Реализация обогащения метаданных:

- Дополните неструктурированные данные описательными метаданными.
- Эти метаданные должны включать информацию об источнике данных, цели, дате создания и авторах.

Использование NLP для извлечения связей:

- Применяйте методы обработки естественного языка (NLP), чтобы автоматически выявлять и извлекать связи между сущностями в ваших данных.

Построение графа знаний с использованием метаданных и связей:

- Используйте обогащённые метаданные и извлечённые связи для построения графа знаний. Начните с малого и расширяйтесь.

Централизация

🧩 Что это значит:

Собрать всю информацию в одном месте, чтобы у AI была доступна единая, проверенная версия данных.

📌 Пример:

Инструкции хранятся в одном Google Docs, FAQ — в другой системе, а полезные советы – в письмах у сотрудников. GenAI не сможет «понять», где правда.

✅ Что делать:

- Перенести всю информацию в централизованную систему (например, Help Center, Notion, Confluence)
- Назначить владельцев разделов, кто отвечает за обновление



Управление промптами

Что это значит:

Создать готовые шаблоны запросов (промптов) для AI, чтобы он отвечал правильно и стабильно.

Пример:

Вместо свободной команды "расскажи, как вернуть товар", используется шаблон:

"Ответь кратко и понятно, с шагами, что нужно сделать, чтобы клиент вернул товар."

Что делать:

- Сделать библиотеку промптов по типам обращений
- Тестировать их на примерах
- Упростить — чтобы любой сотрудник мог использовать



Внедрение RAG (Retrieval-Augmented Generation)

   Что это значит:

GenAI будет **сначала искать нужные документы**, а потом формировать ответ — это как "AI с доступом к базе".

 **Пример:**

Клиент спрашивает: "Как сменить тариф?"

AI ищет инструкцию → находит → вставляет в ответ понятное объяснение.

 **Что делать:**

- Подготовить базу знаний для поиска (разбить на части, описать тематику)
- Настроить векторный поиск (например, через Elasticsearch, Weaviate)
- Проверить, что AI действительно ссылается на нужные документы

ugmentation



MENTED G

Обеспечение безопасности

🔒 **Что это значит:**

Убедиться, что AI не выдаёт конфиденциальные данные и не нарушает законы (например, GDPR).

📌 **Пример:**

Если AI может "вытянуть" номер паспорта из старого чата и выдать его в ответ — это серьёзное нарушение.

✅ **Что делать:**

- Шифровать данные
- Установить права доступа
- Удалять/маскировать персональные данные
- Провести аудит на соответствие требованиям (например, ISO, GDPR)



Контролируйте выходы GenAI с помощью фильтров качества

Фильтры контента

- Настройте соответствующие фильтры, чтобы система GenAI получала доступ только к нужным хранилищам контента.

Фильтр приватного контента

- Установите фильтр, предотвращающий утечку внутренней или конфиденциальной информации за пределы организации.

Фильтр токсичного и предвзятого контента

- Удаляйте контент, содержащий потенциально токсичную или предвзятую информацию, чтобы избежать вредоносных ответов.

Фильтр дублирующегося контента

- Исключайте избыточную информацию для повышения эффективности и предотвращения перегрузки контекстного окна.

Человек-в-цепочке

 +  **Что это значит:**

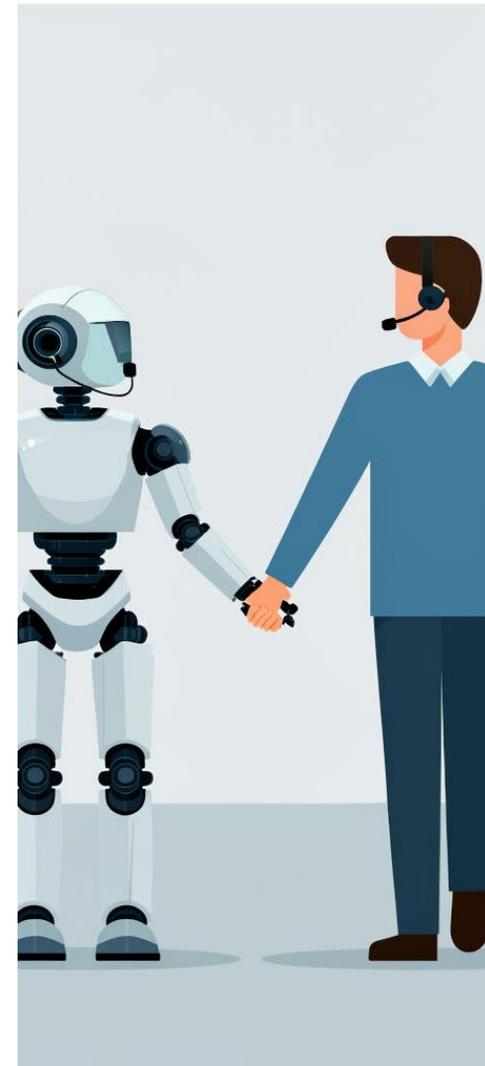
Ответы AI должны проходить проверку человеком — особенно в начале.

 **Пример:**

AI предлагает новый текст в шаблон письма — специалист проверяет, правит при необходимости и одобряет.

 **Что делать:**

- Настроить процесс предварительного одобрения
- Выделить «контент-редакторов»
- Постепенно снижать ручную проверку, когда качество стабилизируется



Борьба с предвзятостью



Что это значит:

Следить, чтобы AI не «заучивал» и не повторял предвзятые суждения, ошибки или устаревшие практики.



Пример:

AI предлагает клиенту "обратиться в офис", хотя компания давно перешла в онлайн.



Что делать:

- Использовать материалы из разных источников (внутренних и внешних)
- Периодически тестировать AI на шаблонные вопросы
- Исключать устаревшие данные и сленг



Изменения культуры организации

 Что это значит:

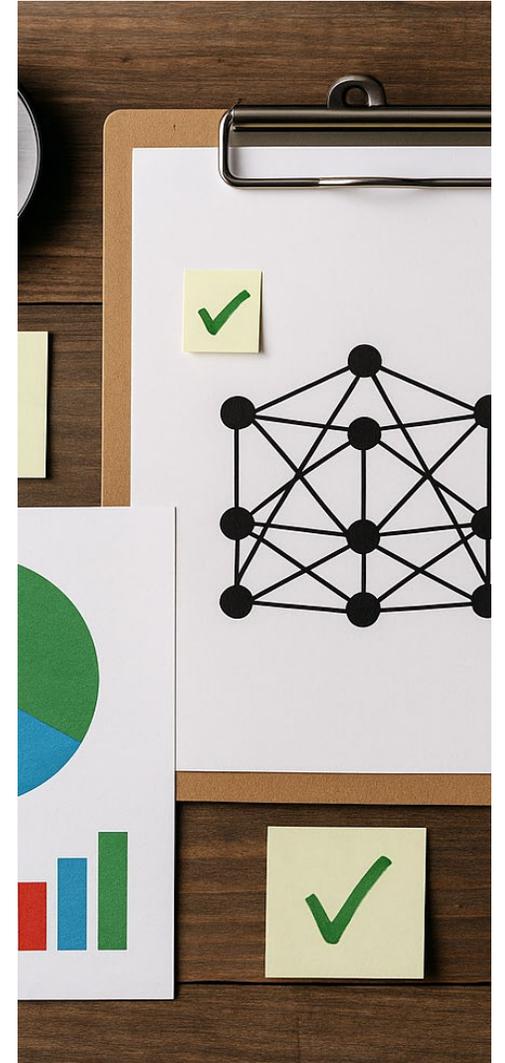
Внедрение GenAI требует изменения мышления: сотрудники должны **доверять AI**, понимать его **помощь**, а не воспринимать как угрозу.

 Пример:

Операторы не хотят использовать AI, потому что боятся, что он их заменит. В итоге — сопротивление, ошибки, саботаж.

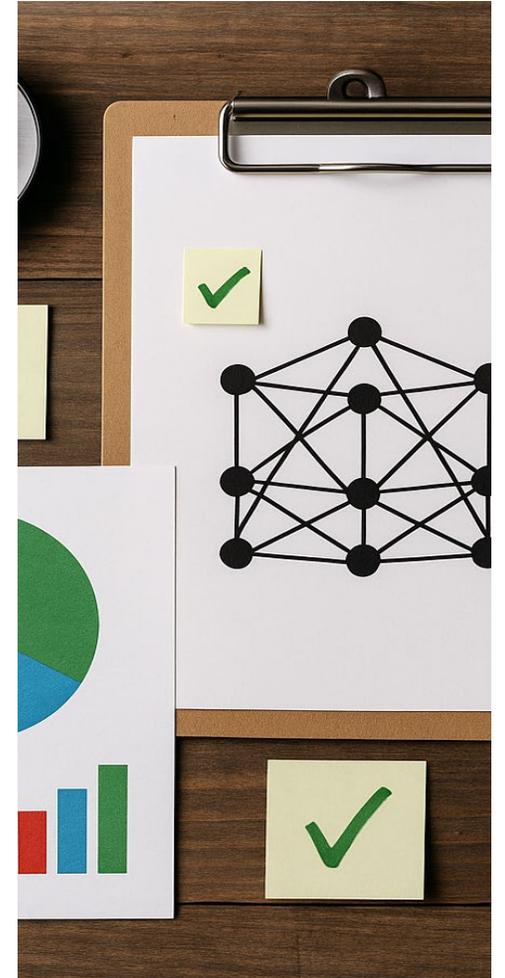
 Что делать:

- Объяснить, зачем нужен GenAI и как он помогает
- Показать успешные кейсы
- Внедрить обучение, игровые задания
- Поощрять сотрудников, кто делится знаниями и улучшает базу



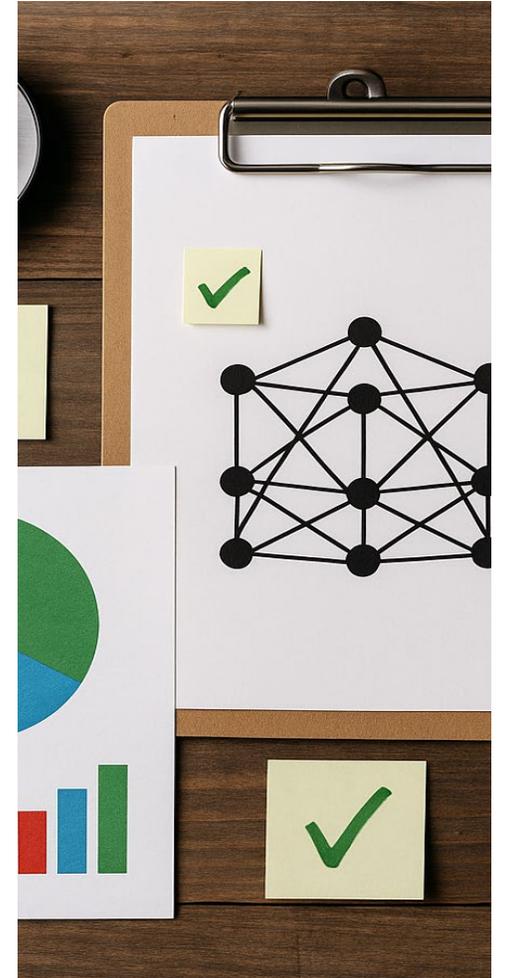
Инвестируйте в современную систему управления знаниями (Knowledge Management)

- Убедитесь, что ваша система управления знаниями поддерживает работу с Генеративным ИИ — имеет готовые коннекторы, открытые API, а также встроенные функции для управления промптом, отбора контента, настройки бизнес-правил и контроля качества.



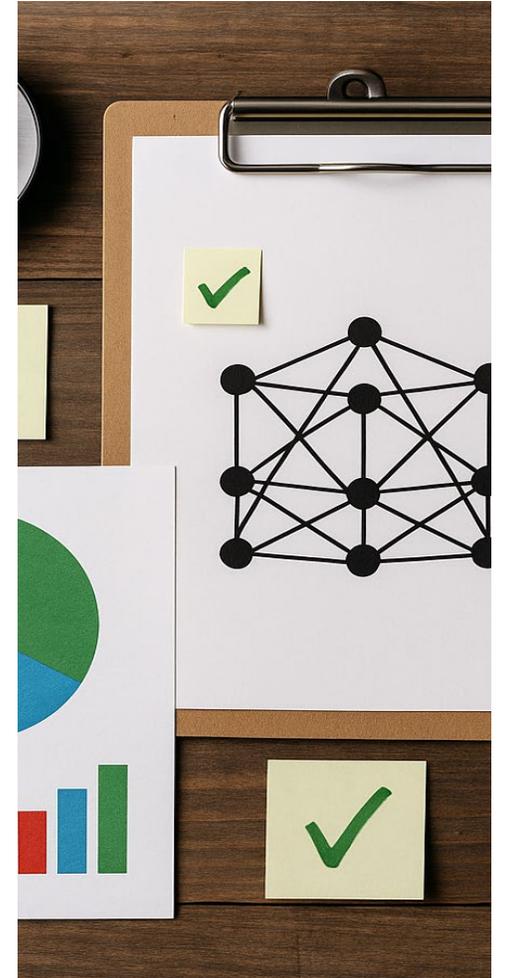
Исключайте разрозненность контента и процессов

- Не допускайте, чтобы GenAI был оторван от основной информационной среды. Интеграция с корпоративной базой знаний критически важна для того, чтобы ИИ получал и использовал достоверную, актуальную информацию.



С самого начала измеряйте всё и управляйте данными (Measure & Manage from Day Zero)

- Если вы не можете что-то измерить — вы не сможете это улучшить. Многие проекты GenAI проваливались именно из-за отсутствия детальной видимости, как ведут себя их AI-инструменты. Внедряйте глубокую аналитику (metrics, KPI) до включения ИИ в продакшен, чтобы видеть динамику «до» и «после» и оптимизировать систему на всех этапах.



СПАСИБО ЗА ВНИМАНИЕ!

apexberg.ru

ТГ-КАНАЛ:
Клиентский сервис –
искусство служить людям

