



ИИ-помощник для сотрудников сети отделений



Леонид Курашов

Лидер кластера чат-боты
ВТБ Онлайн



Леонид Курашов

Лидер кластера
чат-боты ВТБ Онлайн

20 лет

в банках и
финтехе

6 лет

в команде
чат-бота ВТБ

3

кросс-функциональные
команды разработки

Чат-бот ВТБ - абсолютный чемпион



Naumen

Топ 1



Frank RG

Топ 1



Markswebb

Топ 1

Чат-бот ВТБ Онлайн

NLU* и Статические сценарии

Чат-бот распознает намерение клиента и определяет тематику. Темы, в которых чат-бот отвечает заготовленной фразой — статические.

2000 тематик

45% трафика

Интеграционные сценарии

Чат-бот используют данные и сервисы других систем банка для предоставления персонализированной информации или end-to-end сервиса внутри диалога.

140 тематик

55% трафика

Сценарии на базе ИИ

Сценарии, где ответ формируется генеративной моделью на основе источника данных (RAG + LLM) или в свободном диалоге с клиентом, ИИ-агенты

MVP — Чат-бот для сотрудников 2025

Запуск в чат-боте для клиентов — 2026

*NLU — natural language understanding, RAG — Retrieval-Augmented Generation.
LLM — Large Language Model — большая языковая модель.

Боли сотрудников при обслуживании клиентов

01

Много источников для поиска информации, большой объем данных. Сложный и длительный ручной поиск по базе знаний

02

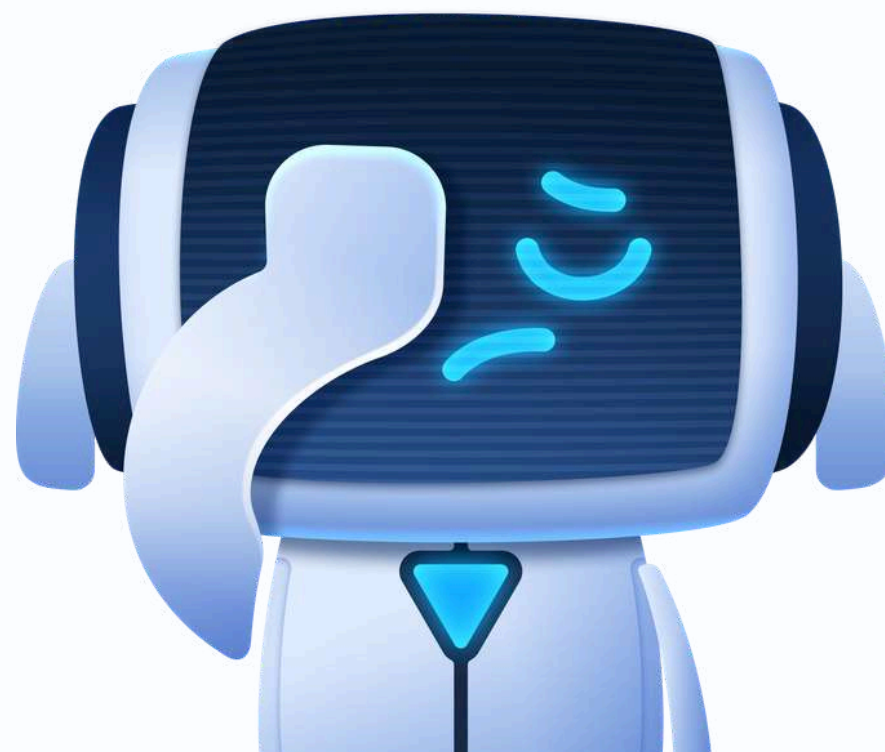
Проблемы с оперативным получением информации об изменениях в банковских продуктах и услугах

03

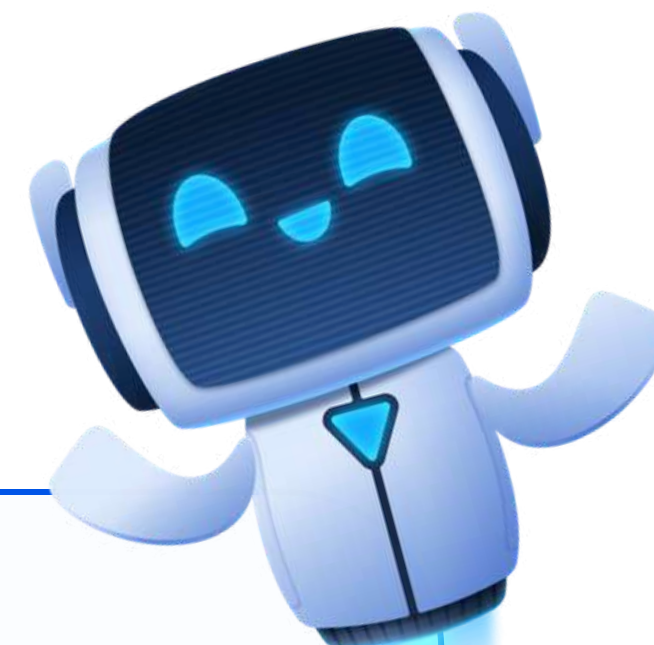
Высокая нагрузка на службу методологической поддержки. Длительное время ожидания ответа на вопросы

04

Неравномерная квалификация сотрудников. Низкое качество обслуживания из-за стресса и неуверенности в ответах



Внедрение ИИ-помощника



Цель

Повысить качество клиентского обслуживания в отделениях банка за счёт обеспечения сотрудников мгновенным доступом к точной, актуальной и унифицированной информации



Решение

Интеллектуальный чат-бот с ИИ, который извлекает релевантные данные из корпоративных источников и формирует точные ответы в режиме реального времени

Состав участников

01

Мессенджеры
и чат-боты

02

Цифровые
агенты

03

Источники
данных

04

Внутрибанковские
продукты

05

Моделирование
RnD

06

Забота
о клиентах

07

Диалогово-
коммуникационная платформа

08

Платформы и решения
для моделирования

09

Филиальная
сеть



Внедрение ИИ-помощника



Этап 1

Сентябрь 2024

Чат-бот для
операционистов



Этап 2

Март 2025

Пилот по ИИ в чат-боте
для операционистов



Этап 3

Октябрь 2025

Масштабирование
ИИ в чат-боте на всю
сеть отделений



Этап 4

Март 2026

Персонализация
ответов в чат-боте





Сентябрь 2024

Этап 1

Чат-бот для операционистов

Чат-бот для оператористов



Чат-бот основан на сценариях обслуживания по наиболее распространенным запросам. Позволяет сотрудникам быстрее находить нужную информацию и стандартизировать ответы

22 тыс.

аудитория
пользователей

6

операторов
поддержки

1100

сценариев
разработано

38 тыс.

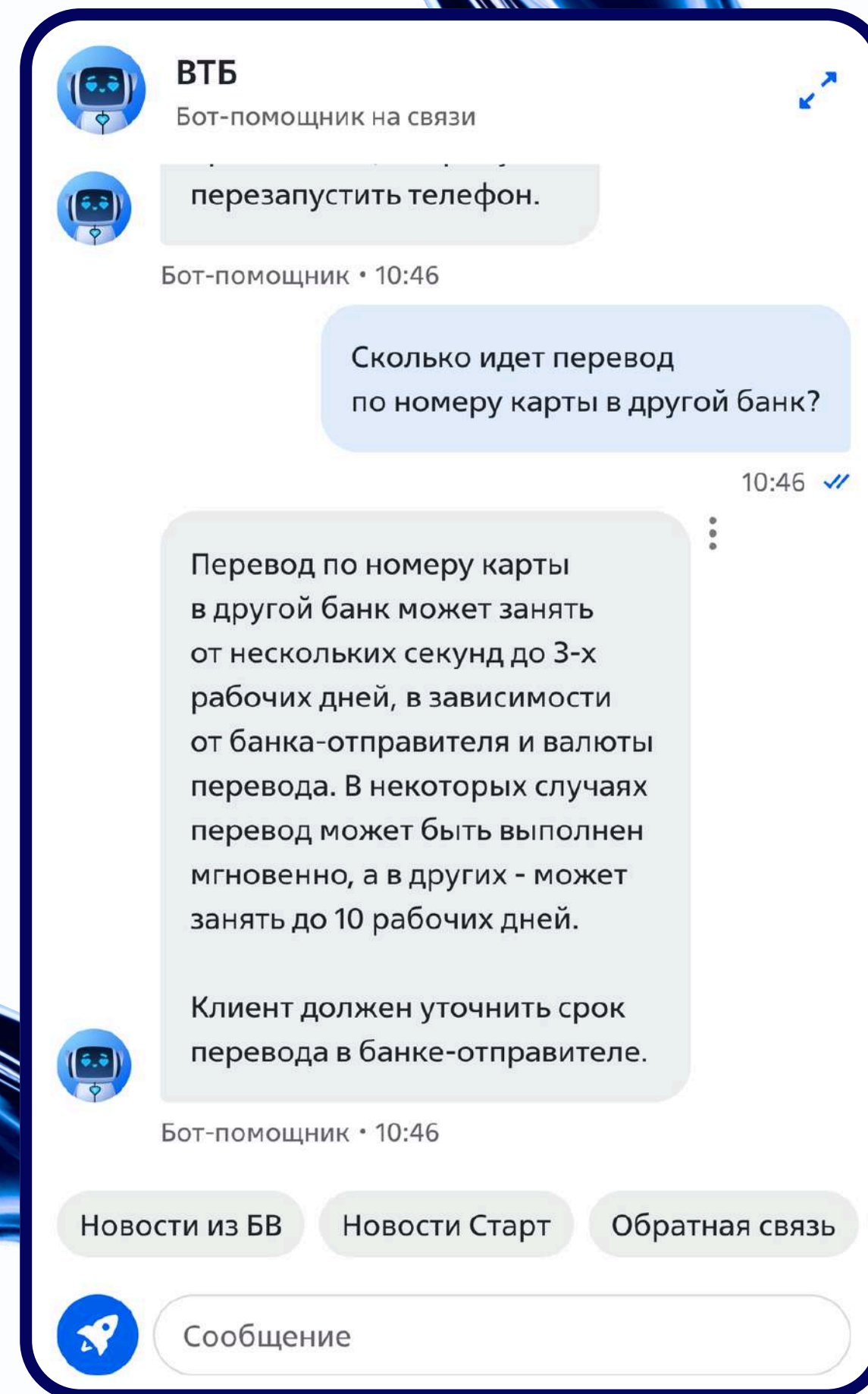
обращений
в месяц

72%

автоматизация
чат-бота

95%

удовлетворенность
ответами бота



Результаты 1 этапа



Проблема

Чат-бот может оказаться невостребованным для сотрудников отделений, так как есть привычный паттерн использовать существующий инструмент — поиск по базе знаний

Решение

Чат-бот оказался настолько удобен, что стал одним из основных инструментов поддержки сотрудников



Проблема

Отсутствие реальных запросов в чат-бот снизит качество ответов модели, так как реальные запросы могут отличаться от выдуманных

Решение

Получили базу реальных запросов в чат-бот, которую взяли за основу для разработки ИИ-сервиса и оценки качества RAG-сервиса





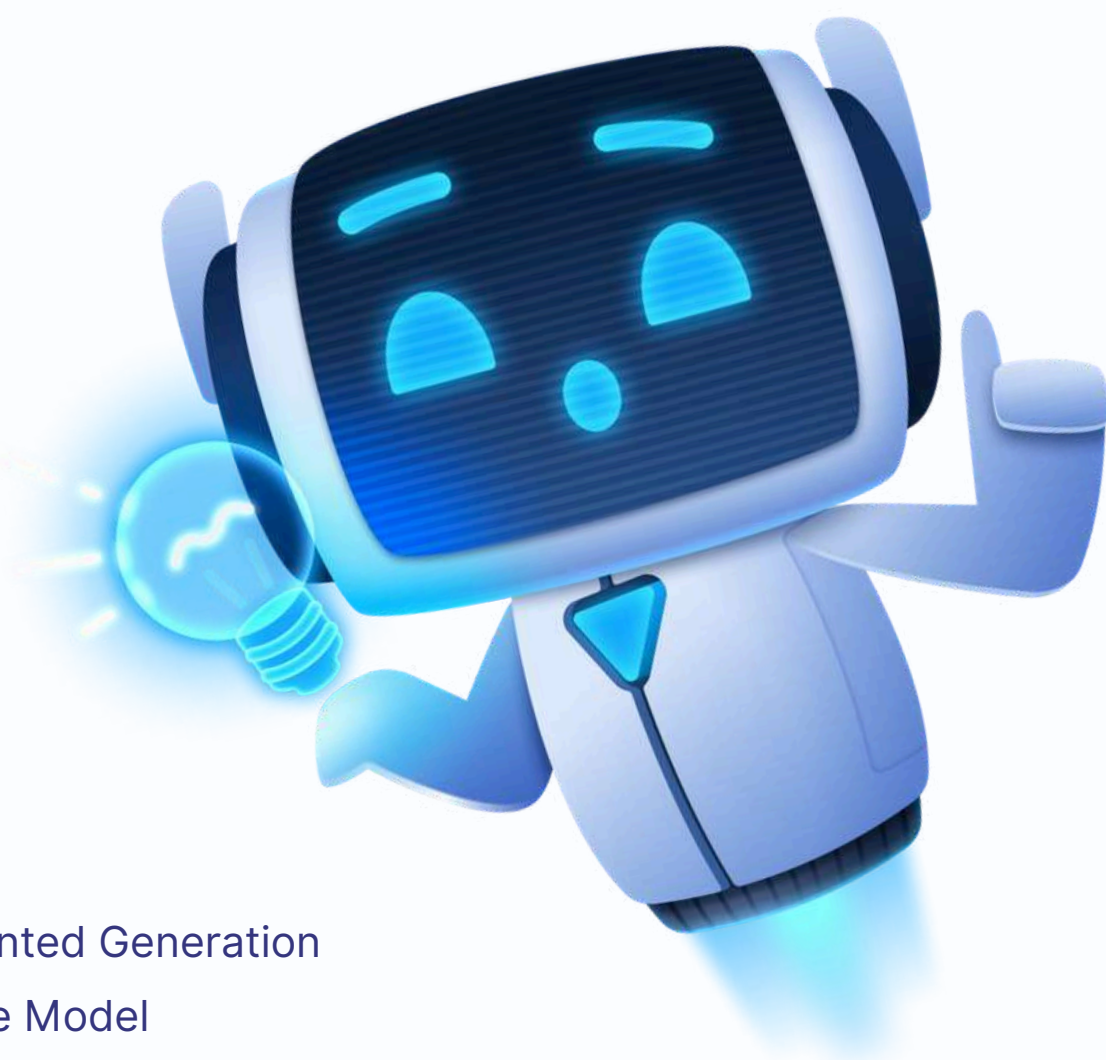
Март 2025

Этап 2

Пилот генеративного
ИИ в чат-боте

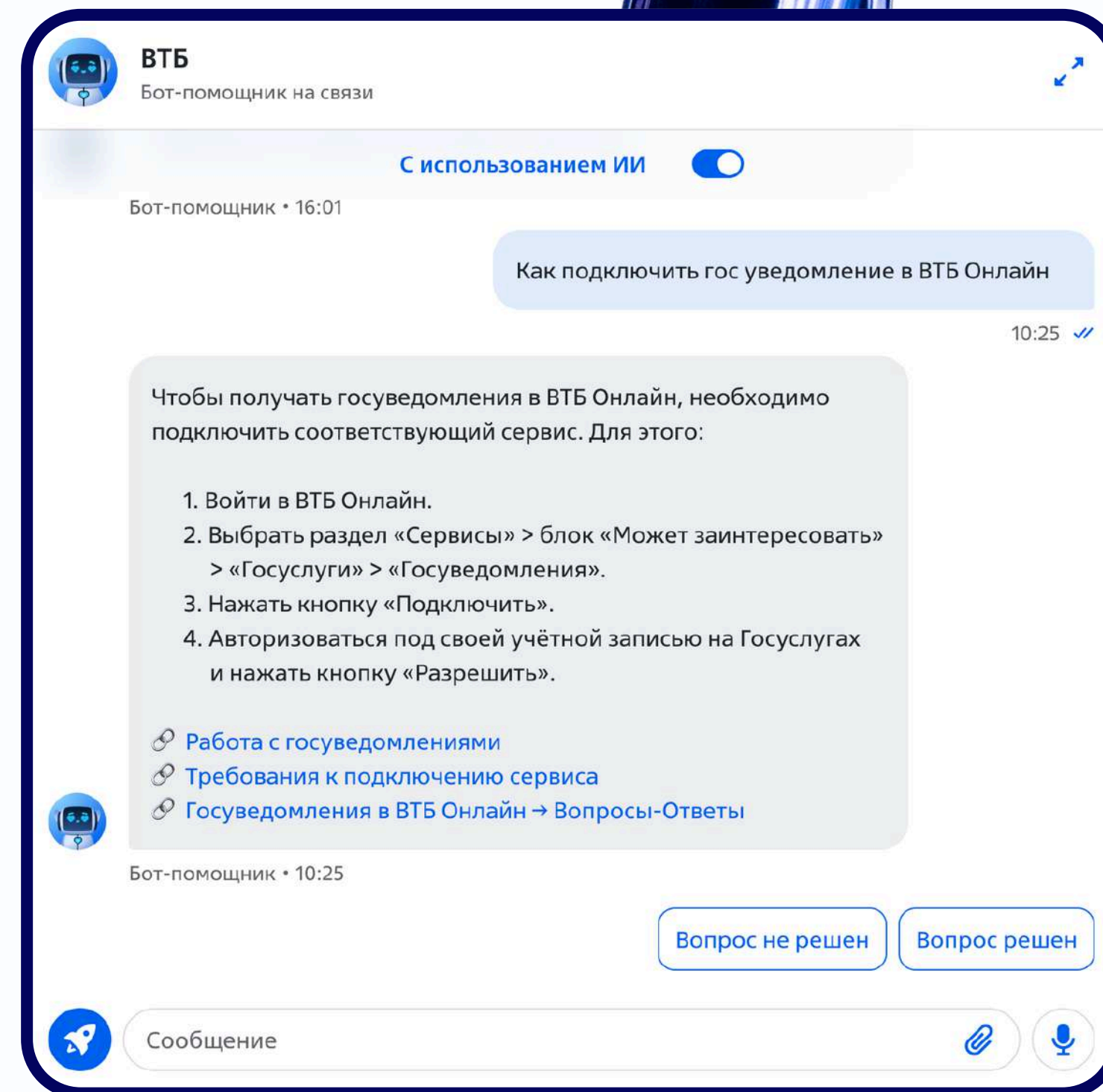
Пилот по ИИ в чат-боте для операционистов

Интеллектуальный чат-бот использует передовые технологии искусственного интеллекта: поисково-дополненную генерацию (RAG*) и большие языковые модели (LLM**)



*RAG — Retrieval Augmented Generation

**LLM — Large Language Model



Что такое RAG



Преимущества RAG



01

Точность и полнота ответов

Высокое качество ответов
за счёт генерации на основе
проверенной и актуальной
информации

02

Адаптивность

Для ответа используются
только актуальные данные
с учетом обновления
базы знаний

03

Оперативность

Быстрое предоставление
релевантных решений
без длительного поиска
и переключения между
источниками

Метрики качества

Результат пилота

9 сек.

Среднее время
ответа сервиса

5 сек.

Целевой показатель

81%

Точность ответов
модели

от **90%**

Целевой показатель



Вызовы

Проблемы



Сотрудники привыкли общаться с ботом короткими фразами и ключевыми словами. Для RAG же нужен подробный запрос



В запросах часто используются банковские термины — модель их не понимает или считает опечатками



Ответы разнородные, не всегда имеют четкую структуру



Ограничение в ресурсах для оценки качества

Решение



Используем гибридную модель: RAG для полных запросов + сценарные ответы для коротких запросов



Добавили справочник банковских терминов и аббревиатур



Добавили Tone of Voice и классификатор запросов с отдельными промптами



Пакетный прогон запросов чат-бота и модель-критик для оценки ответов RAG



Результаты 2 этапа

Проверка гипотез

Гипотеза 1

Сервис с применением RAG можно использовать в контуре банка для ответов на вопросы сотрудников

Подтвердилась

Сотрудники могут использовать RAG-сервис для поиска ответа в базе знаний. Получение ответа от RAG-сервиса **в 10 раз быстрее**, чем самостоятельный поиск информации



Гипотеза 2

Правила ведения страниц, структура и типы вложений на страницах базы знаний подходят для создания RAG-сервиса

Подтвердилась

Существенных изменений правил работы с контентом базы знаний не требуется, страницы не перегружены информацией. Нужно избавиться от архивных вложений

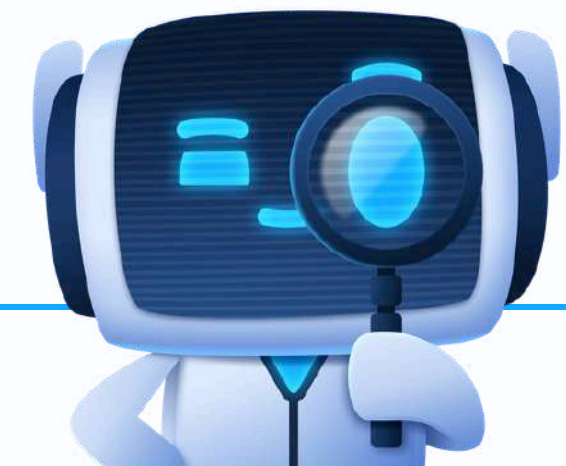


Гипотеза 3

Ответы на вопросы с помощью RAG-сервиса будут короткими, чтобы время генерации укладывалось в 5 секунд

Не подтвердилась

Ответы RAG-сервиса оказались длиннее: среднее время генерации ответа составило **9 секунд**



Октябрь 2025

Этап 3

Масштабирование ИИ
в чат-боте на всю
сеть отделений

Тираж решения



Адаптация виджета чат-бота

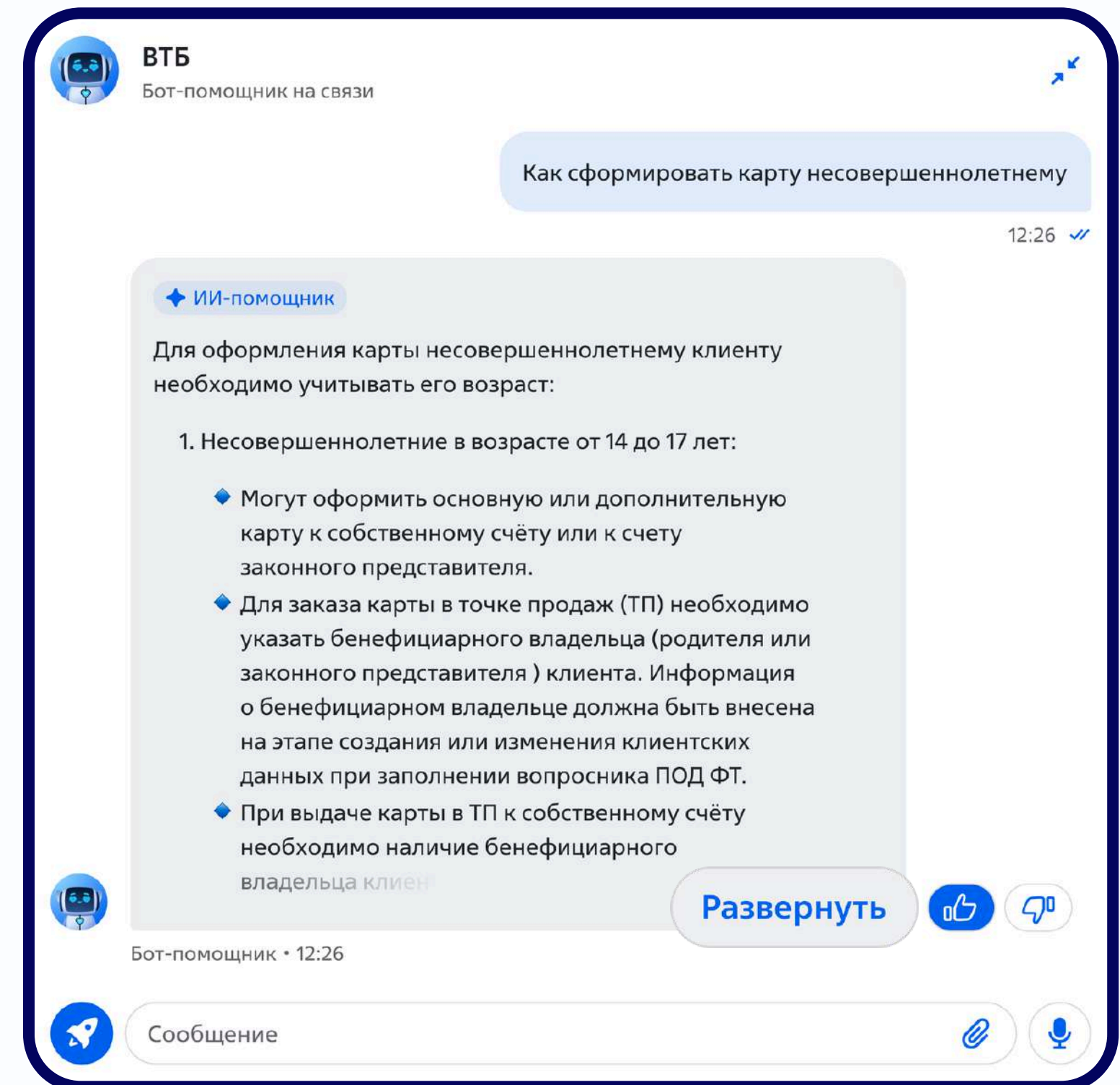
01

Проблема

- ✗ Ответы от RAG-сервиса длиннее сценарных ответов

Решение

- ✓ Расширение размера виджета
- ✓ Скрытие длинных сообщений в истории переписки



Адаптация виджета чат-бота

02

Проблема

- ✗ Необходимость анализировать ответы RAG-сервиса

Решение

- ✓ Сделали видимые реакции возле каждого ответа — лайки и дизлайки

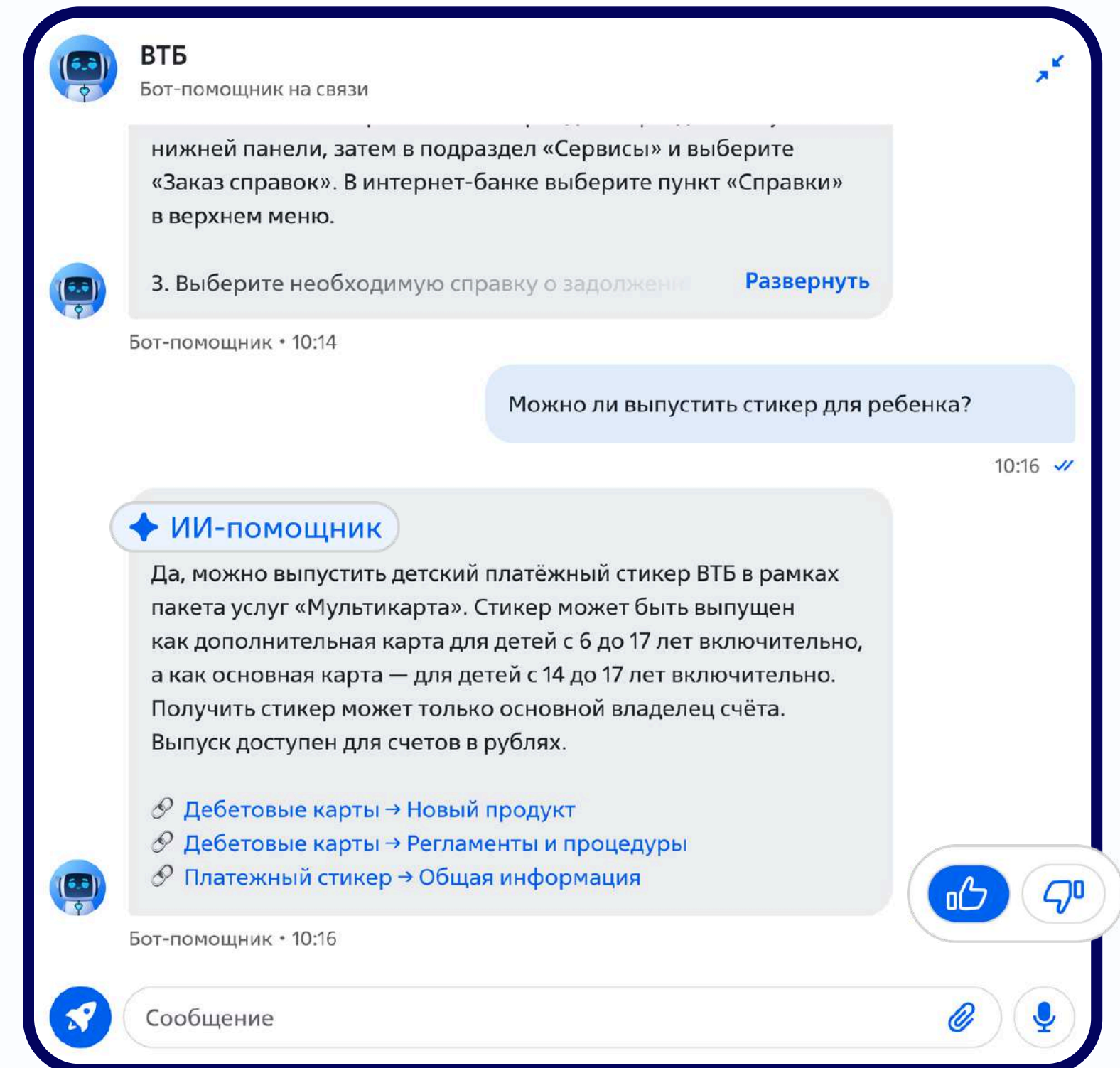
03

Проблема

- ✗ Сотрудники коротко и неполно формулируют запросы

Решение

- ✓ Подсветка ответов RAG-сервиса: «ИИ-помощник»



Результаты 3 этапа

90%

Атоматизация
чат-бота

95%

Удовлетворенность
ответами чат-бота

в 10 раз

Сокращение времени
на поиск информации

80%

Снижение нагрузки
на методологическую
поддержку

в 1,5 раза

Сокращение очередей
в отделениях

75 млн ₽

Экономический эффект
на конец 2025 года

Март 2026

Этап 4

Персонификация
ответов в чат-боте

Планы

Чат-бот будет выдавать персонифицированный ответ



Задача

Интеграция клиентских данных в логику работы ИИ-помощника.



Цель

ИИ-помощник не просто отвечает на вопросы, а предлагает решения для конкретного клиента.

Услуги

Место регистрации

Трудовой стаж

Продукты

Возраст



Персонализация ответов



Наши награды



Национальная банковская премия

За развитие цифровых
и финансовых технологий
в банковском бизнесе

1 место



Data Award

За клиентоцентричность

1 место

Спасибо за внимание!



Леонид Курашов

Лидер кластера чат-боты
ВТБ Онлайн



@LKurashov