

# GIGA B2B

Применение GenAI  
в клиентском сервисе:  
опыт Сбера и Альфа-банка



Лилия Сумбаева

Руководитель группы  
по работе с клиентами,  
СБЕР GIGA B2B



Наталья Балыбердина

Начальник управления развития  
цифровых диалоговых систем, Альфа-  
банк

# Бизнес сегодня созрел для внедрения технологий GenAI

Среди 10 000 крупнейших компаний США и Европы

# 85%

уже внедряют/пилотируют GenAI-инициативы

## Проникновение GenAI, %



# Результаты внедрения ИИ в Сбере



Герман Греф

В 2018 году мы начали масштабную AI-трансформацию Сбербанка, поставив перед собой задачу стать технологической компанией мирового уровня...

Накопленные нами знания, масштаб решаемых задач и полученные результаты могут быть использованы не только в рамках одного нашего банка.

## Результаты внедрения AI в Сбере

- Кредит за 7 минут
- Модель активных продаж (NBA)
- Суфлер в КЦ
- Интеллектуальная система управления
- Чат-боты
- Робот-юрист
- Биометрия
- Фото и видео аналитика

400+  
млрд ₽

Финансовый эффект  
от применения AI  
в Сбере в 2024 году



# GigaChat сегодня. Уникальное сочетание свойств для каждого

**GIGA  
CHAT 2.0**



полностью  
русское решение

ПРОДВИНУТЫЙ  
ФУНКЦИОНАЛ

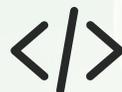
Reasoning

Deep Research

Tools

Вызов внешних  
инструментов

РАЗВИТЫЕ  
МОДАЛЬНОСТИ



ТАМ,  
ГДЕ УДОБНО

Cloud

On-prem

Hybrid



**GIGA** B2B

# GigaB2B — версия нейросети GigaChat, предназначенная для бизнес-заказчиков

Модель обучается на русском языке, под полным контролем процесса претрейна (от сбора данных до профильного обучения)

Отсутствие законодательных рисков — решение состоит в реестре ПО Минцифры РФ

Простая интеграция в бизнес-процессы заказчика в сравнении с open source моделями

Возможность установки on-premise в формате ПАКА



## Российские компании могут:

- Уменьшить выплаты по налогу на прибыль
- Получать субсидии из федерального бюджета
- Участвовать в конкурсе «Лидеры AI» и повышать свой имидж

# Наиболее частотные кейсы применения технологий GenAI

## Co-pilot аналитика

### Аналитика и стратегия

**в 2 раза**  
Выросла скорость  
подготовки отчетности

## Умный поиск по корп.базе знаний

### HR

**в 10 раз**  
Ускорился поиск  
информации

**на 50%**  
Выросла производительность  
труда

## Чат-бот поддержки клиентов

### Клиентский сервис

**на 74%**  
Увеличилось качество  
обслуживания

**в 2 раза**  
Выросла скорость ответа  
клиентам

## Co-pilot для закупщика

### Закупки

**В 10 раз**  
Ниже трудозатраты  
на закупки

**До 20%**  
Дополнительный  
экономический эффект за счет  
экономии на закупках

## Co-pilot для юриста

### Юридическая поддержка

**в 3 раза**  
Сократились трудозатраты юриста на анализ  
правовых источников

## Работа с отзывами

### Клиентский сервис

**в 6 раз**  
Сократилось время отработки  
негативных отзывов

**на 30%**  
Рост CSI

## Генерация контента

### Маркетинг

**в 20 раз быстрее**  
Генерация текстов, статей,  
пресс-релизов

## Анализ маржинального дохода

### Финансы

Повышение точности прогнозирования  
Увеличение прибыли за счет оптимизации  
процессов

# Чем мы можем быть полезны

От консалтинга до доступа к технологии

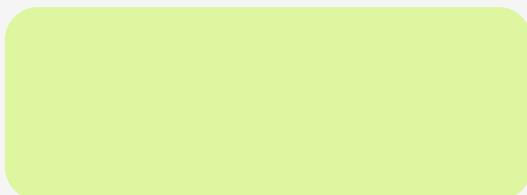


@LILIYASUMBAEVA



## ■ Консалтинг —

комплекс работ по повышению качества ответов модели в различных сценариях применения, проработка и тестирование гипотез, поиск оптимальных пайплайнов обработки информации и реализация целевых сценариев



## ■ AgentNavigator —

Low-code платформа для создания цифровых помощников сотрудников, обладающих искусственным интеллектом, навыками и способностью самостоятельно работать в информационных системах клиента



## ■ GigaB2B —

Интегрированная нейросеть GigaChat 2.0 в бизнес заказчика, вариант поставки в зависимости от выбранного варианта, on-prem или cloud



## ■ Инфраструктура

Серверное оборудование или Cloud (CPU, GPU, RAM, HDD, Сеть)

# Клиентский кейс: Альфа-банк

# ЭВОЛЮЦИЯ AI В БОТАХ

От NLU к LLM



## СЕЙЧАС

## ВЫЗОВЫ

## КУДА ИДЕМ?

### ТЕХНОЛОГИЧЕСКИЙ ФУНДАМЕНТ



Для создания любого бота

Бот-система

Распознавание речи

NLU-модели



JAICP



### ВНУТРЕННИЙ ЦЕНТР КОМПЕТЕНЦИЙ

- Развитие систем и Data Science
- Сопровождение систем
- ЦК RUN-сценаристов и CX-специалистов
- Разметчики данных



### 1 Оценки ниже, чем у человека

*Предвзятость к NLU-боту*

### 2 Рост доли сложных вопросов

*Требуется еще больше переменных данных и их интерпретация, как у естественного интеллекта*

### 3 Неестественность общения

*Нет эмпатии и искусства ведения свободного диалога*



### Альфа-Банк:

- Copilot ассистент оператора
- AI-ассистенты и агенты с LLM
- Суммаризация в базе знаний
- RAG

### Рынок:

- Фокус на RAG. Создание Copilot-систем, где человек и ИИ дополняют друг друга
- Создание небольших AI-агентов, которые принимают решения за человека и подсказывают человеку

# Альфа-Помощники

Семейство сервисных AI-ассистентов

A

с 2019 г.  
**Chat-Bot**

с 2023 г.  
**AI-Ассистент  
оператора**

с 2024 г.  
**Виртуальный  
оператор КЦ  
(LLM)**

2025 г.  
**AI-agents**

# Роль чат-ботов в потоке обращений

Активная клиентская база

100%  
ПОТОКА

Сегментация

Чат-бот

79.8%

Подходит под автоматизацию

20.2% Трансфер до бота:

- Не завершил диалог с оператором
- Продажи и удержание
- Сразу попросил оператора

Оператор

28.8%

Трансфер после бота:

- Клиент попросил оператора
- Модель не уверена
- Требуется расследование
- Прочие трансферы

51%

АВТОМАТИЗАЦИЯ 100% ПОТОКА

Чистая эффективность бота  
**63.9%**



СЛОЖНЫЕ  
ТЕМАТИКИ



ПРОСТЫЕ  
ТЕМАТИКИ

# ЧАТ-БОТ НА NLU

«ЛУЧШИЙ ЧАТ-БОТ» 3 года подряд в рейтинге Chat-Bot Rank Markswebb



## 90% ПОКРЫТО ИНТЕГРАЦИЯМИ

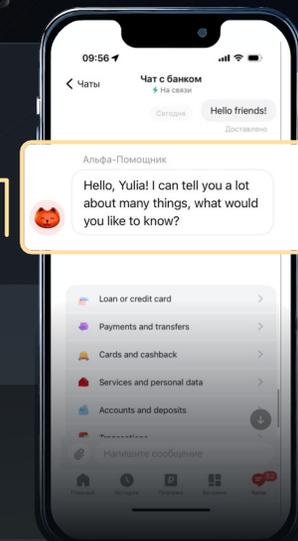
- ✓ Модернизация с акцентом на автоматизацию
- ✓ Глубокая переработка сценариев и интерфейсов
- ✓ Создание целевых интеграций, в т.ч. для LLM

## ВНЕДРЕНИЕ МОДЕЛЕЙ В ЧАТ-БОТ

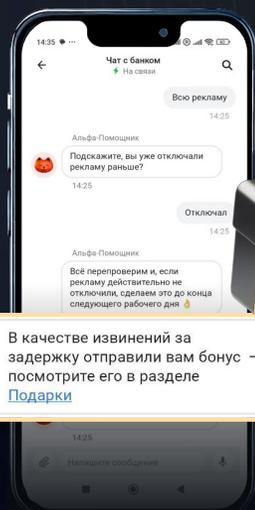
- ✓ Модель рекомендаций по контексту
- ✓ Модель пулов
- ✓ Модель лайков
- ✓ Модель доспрашиваний на LLM

## ПОДДЕРЖКА МУЛЬТИЯЗЫЧНОСТИ

- ✓ Переводчик запущен на 10% пользователей
- Бот разговаривает на всех языках мира



## ПРОГРАММА ЛОЯЛЬНОСТИ



Извинительные подарки и бонусы

Реверсивный cash-back

В качестве извинений за задержку отправили вам бонус – посмотрите его в разделе [Подарки](#)

## ОБУЧАЮЩИЕ КРУЖОЧКИ И АНИМИРОВАННЫЕ СТИКЕРЫ



## ПРЕДИКТИВНЫЕ УВЕДОМЛЕНИЯ



Оповещение клиента

- Об авариях в банке
- О списаниях со счета
- Об обновлении версии приложений

# AI-Ассистент оператора



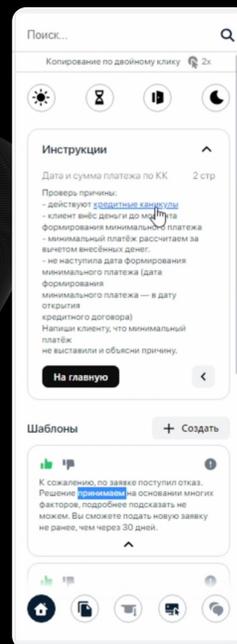
A

Единая  
платформа  
с чат-ботом

Масштабировано  
на 100% в чатах,  
голосе

Оптимизирует  
время работы  
оператора

## Рабочее место оператора КЦ



- 3 120 количество пользователей
- Использует лучшие ответы операторов (NLU золотых ответов)
- 49.5% полезности
- 4.94 VOE
- Интеграция с 6 сервисами LLM



# Что имеем?

## Каскад NLU-моделей

A

### Chat-Bot 14 моделей

Основной классификатор интенгов с выделением сущностей  
**350+ интенгов**

Модели рекомендаций Контекста диалога

Распознавание текста на изображении

А также:  
Модель склонности к низкому VoC  
Модель пулов  
Модель лайков  
Модель саджестов  
И т.д.

Намерение

Чипсы

Сущности



Бот решает проблему

Трансфер на оператора

### Альфа-помощник оператора 3 модели

1  
Рекомендация шаблонов

Все ставки доступны по ссылке [www.alfabank.ru/stavky](http://www.alfabank.ru/stavky)

2  
Рекомендация интерфейсов

Операции клиента

3  
Рекомендация готовых «золотых» ответов

Спасибо, что подождали. К сожалению, по заявке получили отказ банка.

# Что имеем?

A

**200+ интеграций  
в сценариях**

**Центры компетенций:  
разработки, сценариев и СХ**

1

## **Данные нужны любому боту**

Доля сложных тематик самая большая – это всегда расследования, где нужны данные

2

## **Интеграции дают больше автоматизации**

Это большая работа, которую нужно было делать раньше и нужно продолжать делать дальше

3

## **Синергия данных**

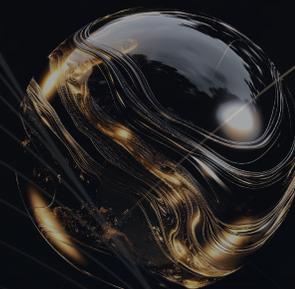
Единая интеграционная шина переменных данных для всех AI-ассистентов

**CHANGE**



**СХ**

**RUN**

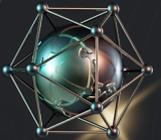


# Синергия поколений ботов



## РАЗДЕЛЕНИЕ ОБЯЗАННОСТЕЙ

NLU чат-бот выполняет простые и сложные задачи, где требуется детерминированный подход, а LLM создает креативный контент для клиента, мимикрируя под оператора



## ОБМЕН ПЕРЕМЕННЫМИ ДАННЫМИ

NLU чат-бот дает переменные данные для ответов, а LLM – уточняет детали вопроса, ведет естественный диалог и заботится о клиенте с кастомным ToV



## БАЛАНС УПРАВЛЕНИЯ АВТОМАТИЗАЦИЕЙ

На NLU оставляем детерминированные сценарии, с поддержкой сценаристов, на LLM переводим сложные вопросы, требующие генерируемый ответ с помощью RAG

# Виртуальный оператор КЦ

Эволюция чат-бота: от NLU к эмпатичному LLM

A

**65 простых интенгов**  
**4 сложных интеграционных интента**

Работает на основе генеративного искусственного интеллекта

Генерация ответов не по шаблону

Использование векторного хранилища в качестве Базы знаний

Вариативность ответов

Поддержка эмпатии

Работа с интеграциями

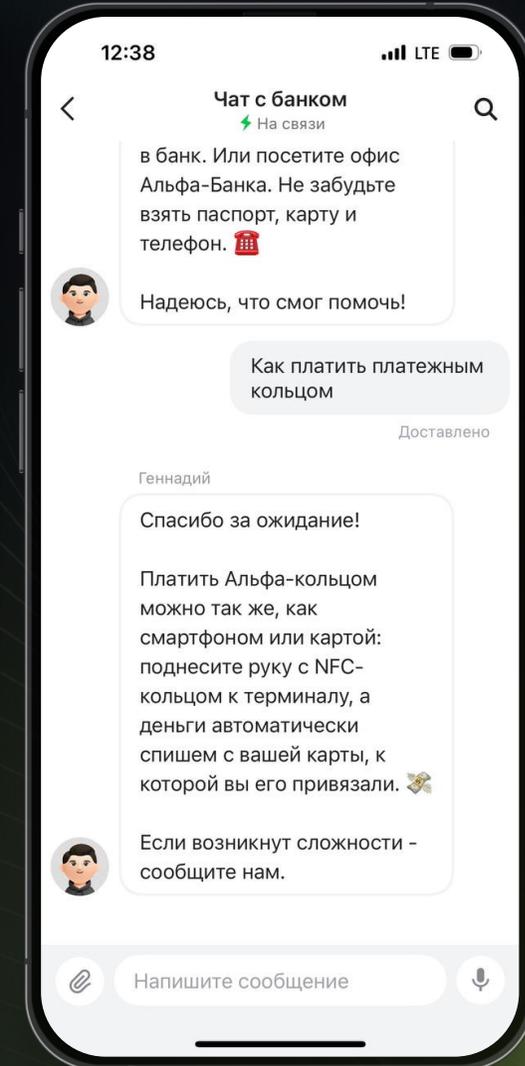
Имитация человеческого общения

- Используются фразы наших лучших операторов
- Применяются задержки ответов для имитации человека
- Подстраивается под стиль общения
- Выявляет эмоциональный окрас запроса пользователя

Геннадий Рагов



Работает на LLM  
Gigachat

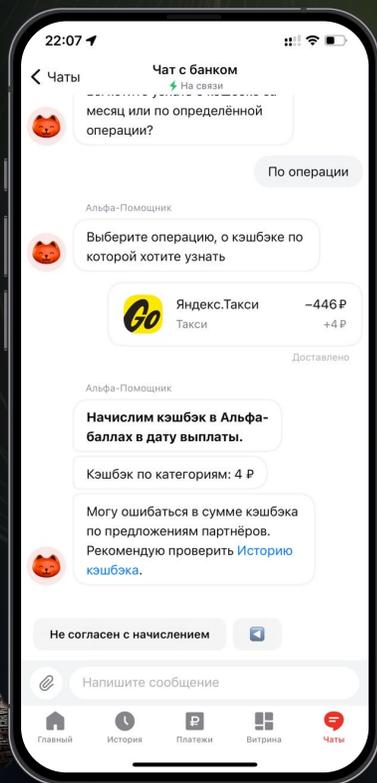




# AI помогает обрабатывать 100% обращений клиентов

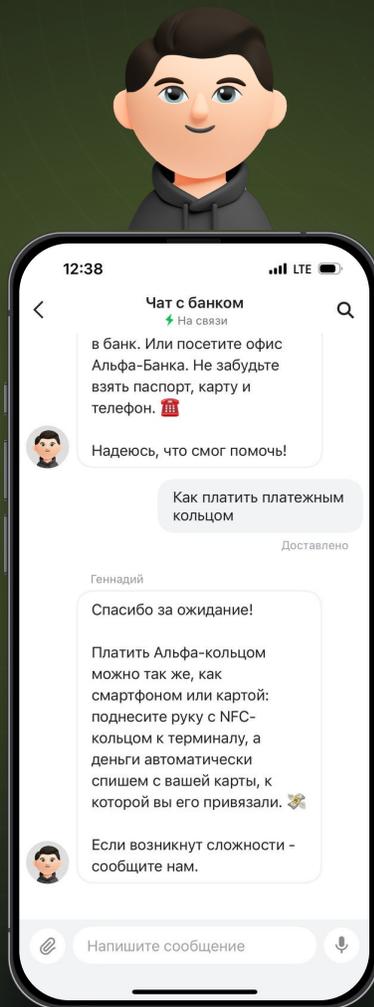
## ЧАТ-БОТ с NLU

Классический бот



## ЧАТ-БОТ с LLM

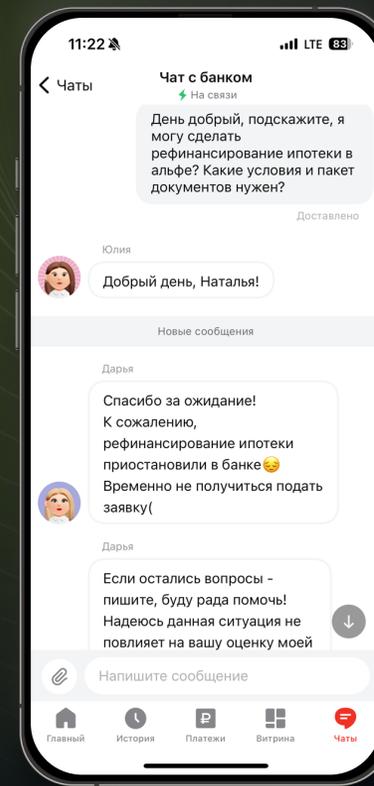
Виртуальный оператор



## AI-Ассистент

оператора

На NLU и LLM





Спасибо за внимание



**НАТАЛЬЯ  
БАЛЫБЕРДИНА**

Начальник управления развития  
цифровых диалоговых систем  
Альфа-Банк

