



**ТИНЬКОФФ**

# **Tinkoff VoiceKit и Voice Conversion: насколько нам принадлежит наш голос**

# Принцип работы VoiceKit на примере голосовых роботов

Человек  
говорит



Речь превращается  
в текст



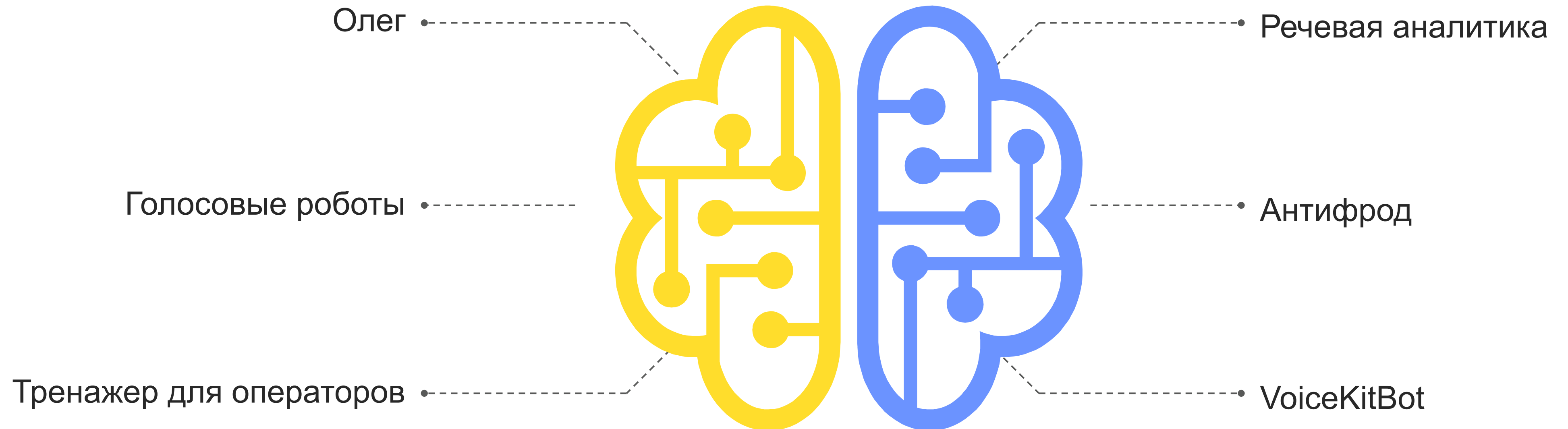
Модель понимает  
смысл и принимает  
решение об ответе



Синтезированный  
ответ озвучивается  
человеку



# Продукты на базе Tinkoff VoiceKit



# Преимущества VoiceKit

## Собственная разработка

- 3 RnD команды
- Большой объем собственных данных
- Суперкомпьютер «Колмогоров»

## Проверено бизнесом

- 40 технологических партнеров
- 150 прямых клиентов

# Показатели качества

## Распознавание

### WER

Tinkoff VoiceKit: 4—12%

- Акценты и диалекты
- Пол и возраст
- Адреса
- Даты и время
- Имена
- Названия и артикулы

## Синтез

### MOS

Tinkoff VoiceKit = 4,3 из 5

- Живость речи
- Время отклика
- Мелодичность звучания
- Интонация
- Корректное произношение имен, сумм, адресов

## Общие

### Latency: задержка распознавания

Tinkoff VoiceKit < 1 с

### SLA

Tinkoff VoiceKit > 99,99



# Какие задачи решают с помощью VoiceKit



## Колл-центры

- Голосовые роботы
- Smart IVR
- Речевая аналитика



## Озвучка текста

- Статьи
- Книги



## Подсказки

- Субтитры
- Суфлеры
- Озвучка алертов



## Интерактив

- Информационные панели
- Игрушки

# Планы



On-premise



Развитие распознавания



Развитие голосовой верификации



Развитие синтеза



**ТИНЬКОФФ**

# Voice Conversion



# Text-to-Speech

Текст → Речь

# Text-to-Speech: синтез речи

Переводит текст в речь в реальном времени. Подходит для голосовых роботов, озвучивания контента и создания брендового голоса.

## Разные голоса

Доступны два мужских и один женский голос: Дорофеев, Максим и Алена

## Брендовый голос

Если нужен персональный голос, помогаем подобрать актера озвучивания и записать набор фраз

## Ударения, паузы и расстановка акцентов

Для большего контроля над синтезированной речью используем SSML: можно добавить паузу, поставить ударение, расставить акценты и настроить посимвольное произношение текста

# Voice Conversion

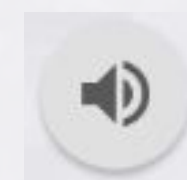
## Speech-to-Speech



target.wav



source.wav




out.wav


Речь → Речь

# Speech-to-Speech: Voice Conversion

## Как применяется технология в контакт-центрах

 Ослабить акцент

 Повысить  
конверсионность  
прозвона

 Деперсонализировать  
оператора

Речь → Речь

# Speech-to-Speech: Voice Conversion

## Как еще применятся технология



Сэкономить  
на работе диктора



Повысить узнаваемость  
бренда



Для озвучки  
в сфере  
развлечений



Конвертировать  
уже записанные  
диктором фразы

# Перспективы развития



Исследование новых возможностей



Сведение к нулю задержки сервиса



Развитие устойчивости к нагрузкам



Формирование библиотеки голосов

# Спасибо

Задавайте вопросы



Отправляйте  
аудиозаписи  
в телеграм-чат



[@voicekit\\_sales@tinkoff.ru](https://t.me/voicekit_sales)