

По дороге с облаками *или* История голоса Альфа



Никита Ткачев
Руководитель развития бизнеса
ML-продуктов Yandex.Cloud



Владимир Китляр
Digital CPO, Альфа-Банк
kitliar.me



Знакомьтесь, Alf

RUN, ALF

RUN

RUN

RUN

Более 1 млн

умных и свободных уже
познакомились с Альфом
за первый месяц



Дилемма инноватора — делать всё самим инхаус?



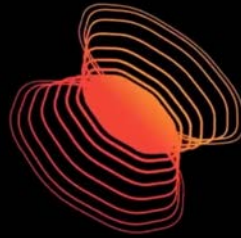
0
0
7

Или довериться профессионалам в тонких R&D-специализациях



**Чтобы сфокусироваться
на специализации и быть
виртуозом своего суперсервиса**





Стратегия эффективных партнёрств с best-in-class-лидерами

Yandex  Cloud

ML-сервисы

ART.LEBEDEV

Знаковые системы

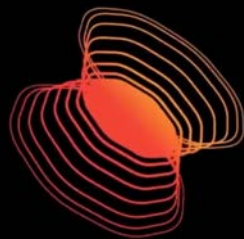


Just AI

AI-технологии,
NLU

 Digital

Развитие платформы



Alf говорит, почти как человек — с помощью облачных сервисов и ML-технологий Яндекса

Уникальный кейс использования облаков в
банковском голосовом помощнике

3 мифа про облака и финансы



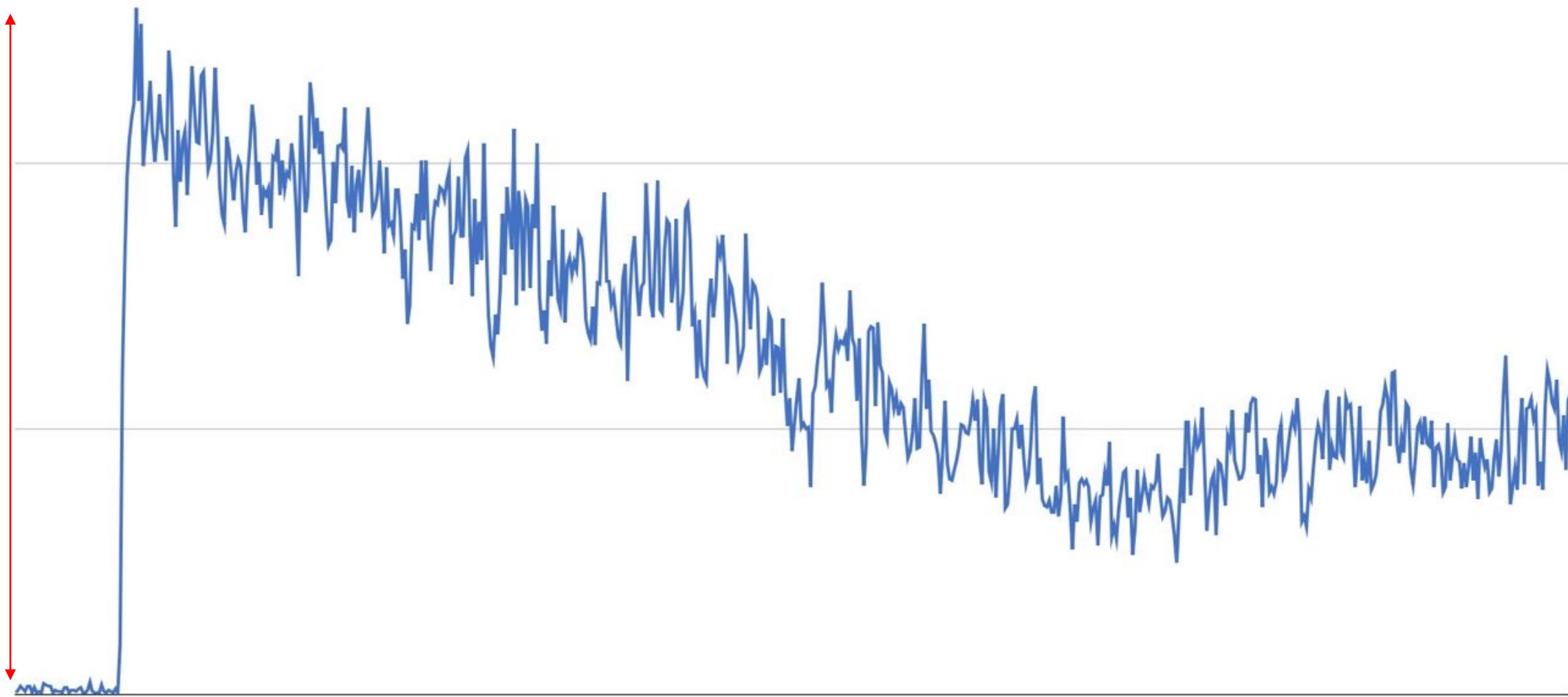
Миф №1



**Облака
не надёжны**

Пример турбулентного трафика

D



Турбулентность трафика R&D-продуктов и спасительные облака

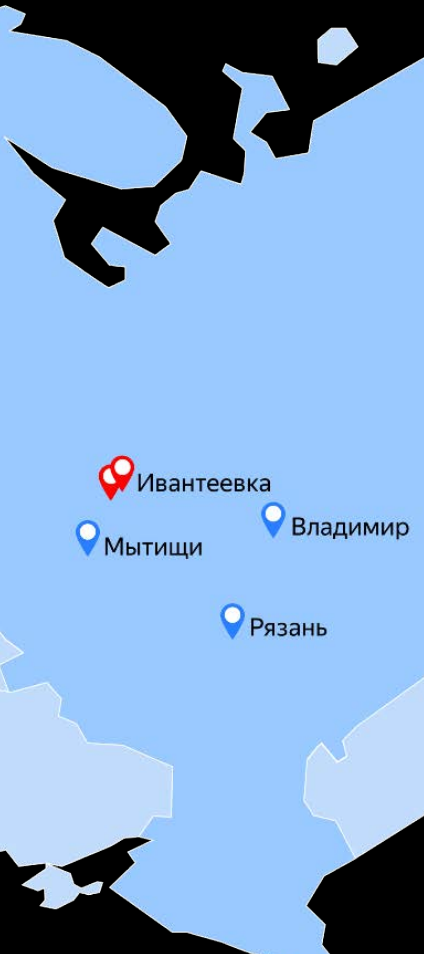


Миф №2



**Облачные
сервисы где-то
за океаном**

Облака, которые можно потрогать



Миф №3



**Облачные
сервисы
негибкие**

BENNY'S
Original
• MOTOR WORKS •

код

хотите покинуть Benny's Original
Motor Works?

**Облачные сервисы не отменяют
возможность индивидуальной
кастомизации под продукт**

D



Под капотом у Альфа или собираем автомобиль на ходу

Никита Ткачев

Из чего состоит персонаж

Портрет

Голос
и визуальное
представление



Речь

Фразы
и манера
диалога



Поступки

Пользовательские
сценарии





Алиса

- Узнаваемый голос персонажа
- Эмоциональная речь богатая интонациями
- Высокие показатели метрик на крауде

Что нужно,
чтобы повторить
и масштабировать
такие проекты?

Три составляющие успеха

Чтобы было «быстро, дешево и хорошо»



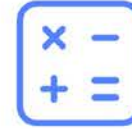
Data

Собрать хорошие
и консистентные данные



Tech

Разобраться какие технологии
лучше всего подходят
для конкретной задачи



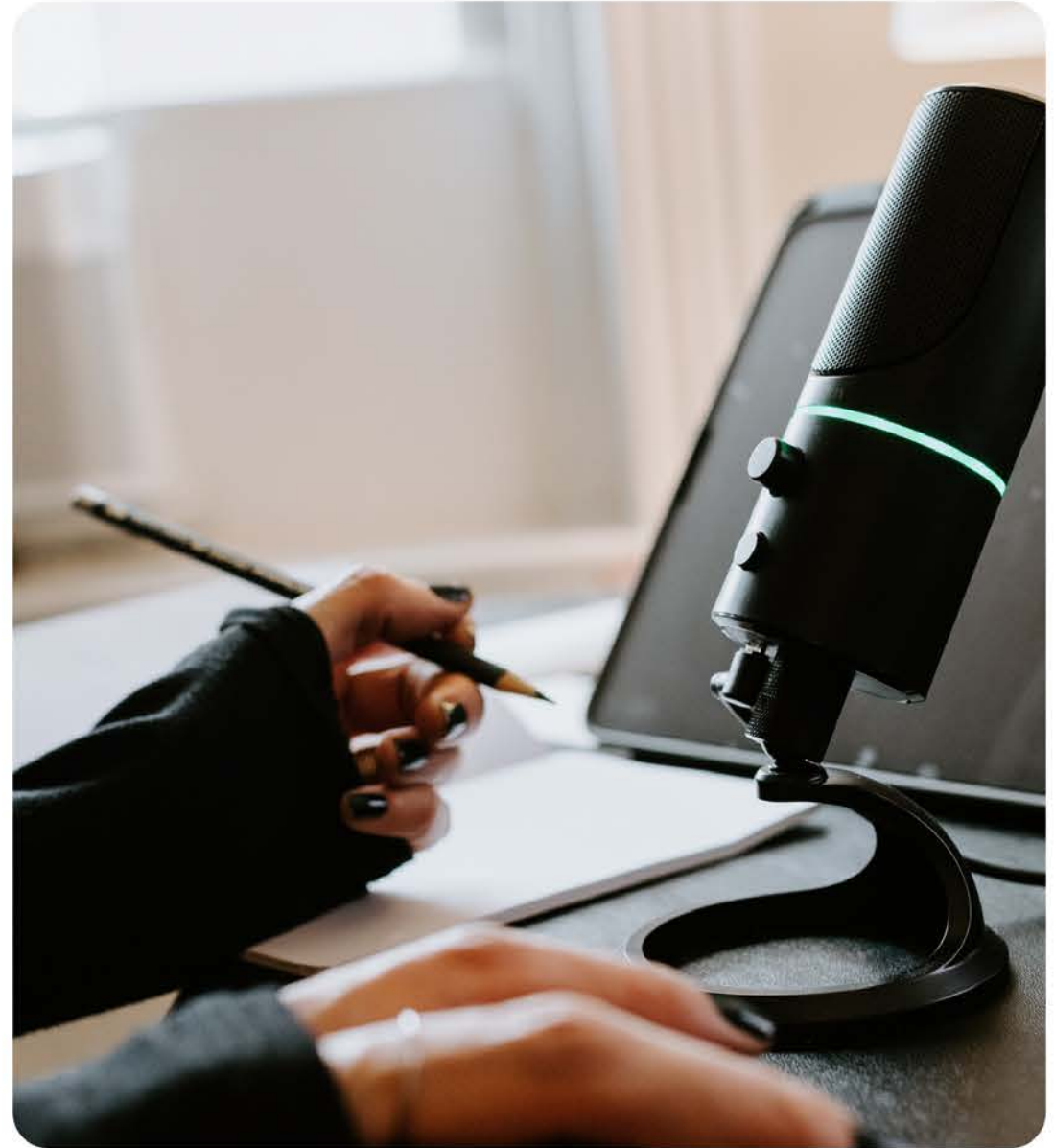
QA

Сделать максимально
объективную оценку
результата

Как мы готовим данные?

Из буквочек в аудио высокого качества

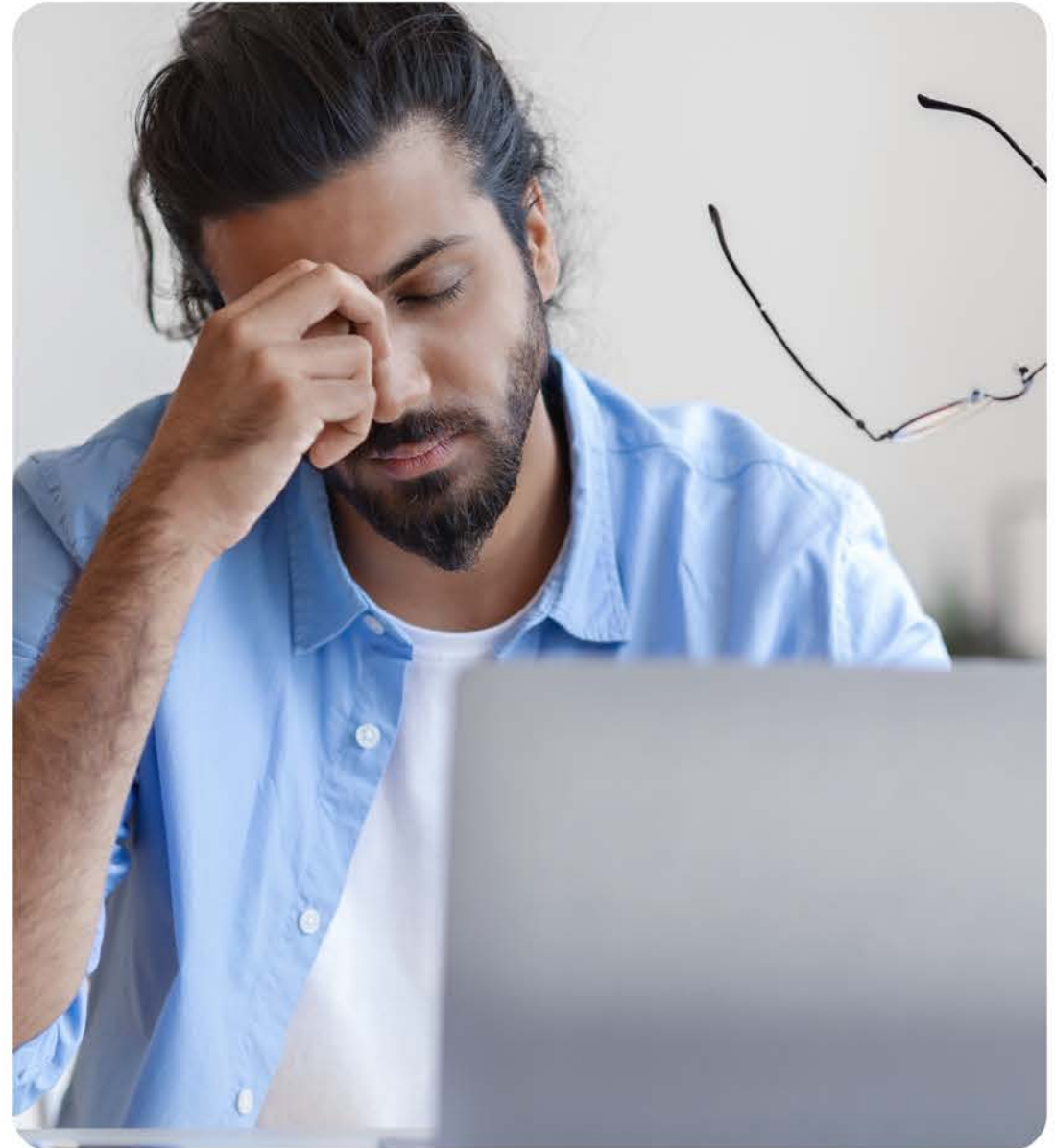
- Генерация текстов под диктора и задачу. Тексты обеспечивают фонетическое разнообразие и соответствуют сценарию. Замешиваем реальные сценарии.
- Собственная команда войс-коучей — составление профиля диктора, контроль консистентности при записи, коррекция диктора, работы студий и текстов.
- Сертифицированные студии и возможность онбординга новых студий, контроль удаленной записи.



Автоматическая валидация датасетов

Что научились автоматически отлавливать

- Ошибки транскрипций
- В студии поменяли микрофон
- 5 лет назад записали кондиционер
- Диктор болел



Технологии

Что мы пробовали и на чем остановились



Акустические модели

- FastSpeech
- Tacotron
- Flowtron
- Vits
- FastPitch

Вокодеры

- WaveGlow
- WaveNet
- HiFi-GAN

Блендинг параметров

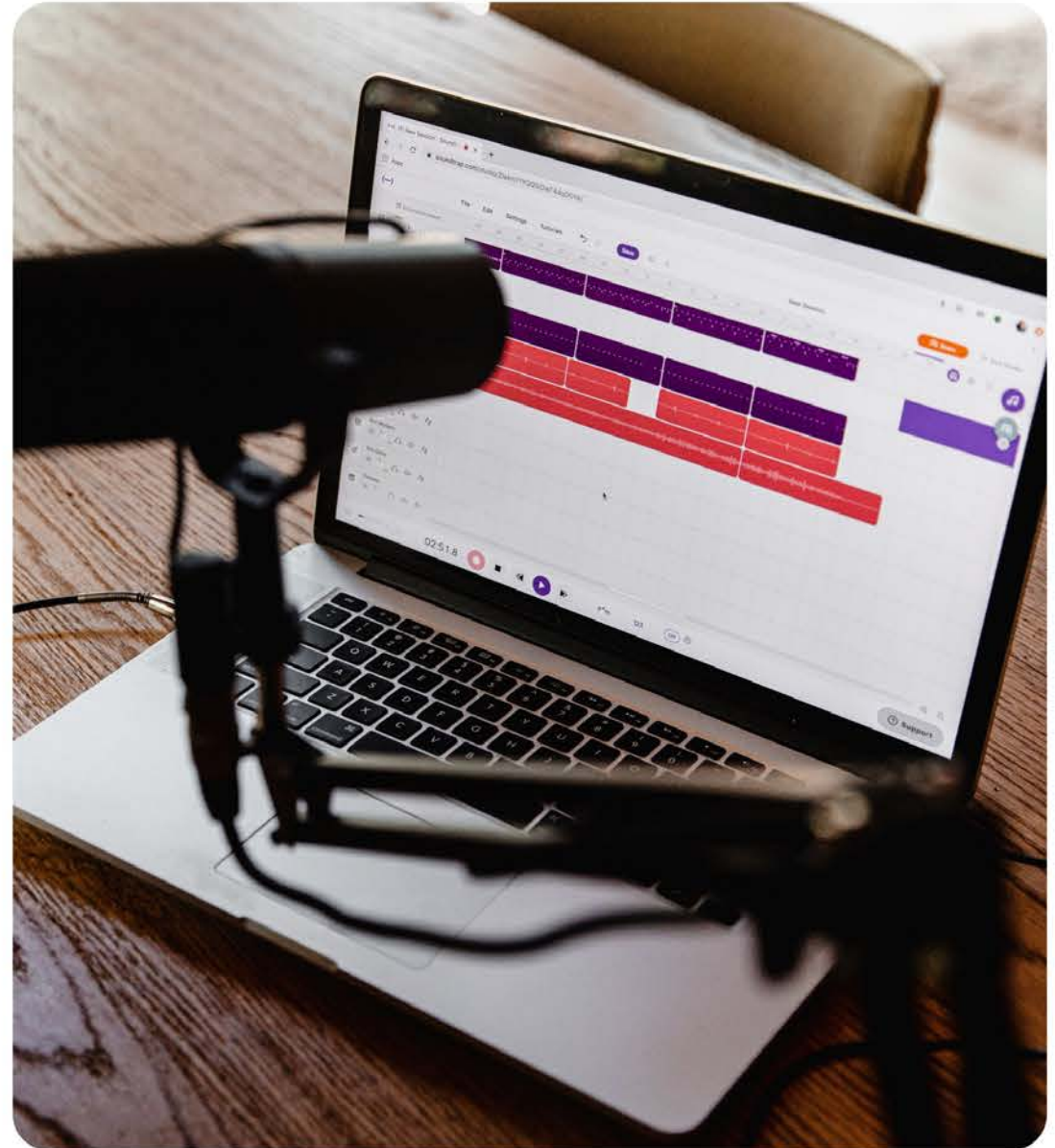
Ничего не делаем и улучшаемся

- Новые модели добавляют знаний о языке
- Готовые модели регулярно ретрейнятся с учетом исправлений и улучшений других моделей
- Полученные модели улучшаются по языку и технике с сохранением голоса актера

Brand Voice Adaptive

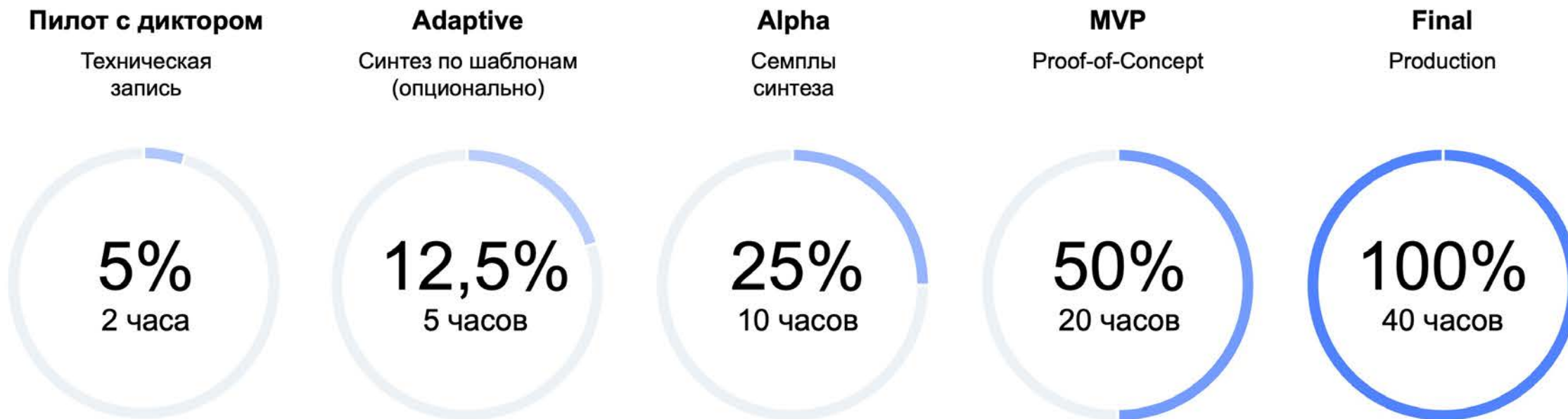
Произвольные интонации

- Наличие датасета позволяет одновременно использовать модель в режиме Adaptive
- В этом режиме помимо текста подается шаблон фразы в виде аудио
- Итоговый синтез копирует интонацию из шаблона с заменами частей текста (адреса, даты и т.д.)



Тренировка

Регулярные итерации, а не кот в мешке



Проблемы с популярными метриками синтеза речи

Side-by-side (SbS)

Сравниваем две итерации модели друг с другом и спрашиваем какая лучше

Проблемы:

- Если не давать текст, то проблемы с вопросительными интонациями не поймать
- Можем улучшить модельку но добавить какой-то новый артефакт (убрали хрип, сломали фонему, но метрика прокрасилась)

Mean opinion score (MOS)

Просим оценить запись по шкале от 1 до 5 и считаем какое-то среднее

Проблемы:

- Попытка натянуть сову на глобус — метрика из телекоммуникаций
- Сильно подвержена «scale bias»
- Нельзя сравнивать средние оценки между собой по разным проектам
- А еще женские голоса в среднем почему-то получают оценки выше мужских ;)

К чему мы пришли?

1

Percent of Indistinguishable Voice Recordings (PIVR)

Предлагаем студию и синтезированную запись (вне трейн-сета), спрашиваем один ли человек на записи или разные.

Метрика бинарная, и мы ее сломали :(

Текущие значения стремятся к 100%

2

Codename MERa*

Набор бинарных метрик по самым распространенным артефактам синтеза (наличие фонового шума, ошибки в интонировании и т.д.). По этому набору метрик подбираются веса, чтобы получить одну итоговую KPI оценку.

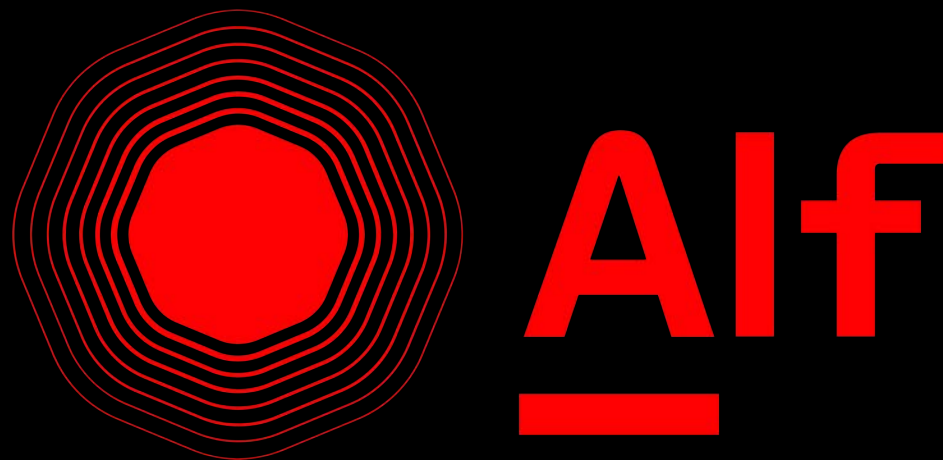


Всеволод Кузнецов
Голос Альфа

D

**Совместная
работа
по кастомизации
синтеза**





Никита Ткачев

Руководитель развития бизнеса ML-
продуктов Yandex.Cloud
niktk@yandex-team.ru

Владимир Китляр

Digital CPO, Альфа-Банк
www.kitliar.me
vkitlyar@alfabank.ru