

# Обзор Bark

## Описание технологии

Bark — это модель на основе трансформеров, которая предназначена для преобразования текста в звук. Она поддерживает многоязычную генерацию речи, невербальную коммуникацию доступ к предварительно обученным контрольным точкам модели для создания высококачественного аудиоконтента. Bark можно использовать для различных задач, связанных со звуком, в том числе для вспомогательных технологий и улучшения технологии преобразования текста в речь.

## Примеры использования

- Создатели контента — для создания уникального и разнообразного аудиоконтента для видео, подкастов и многого другого.
- Звукорежиссеры — для быстрого создания прототипов звуковых эффектов и окружающего звука для различных медиа.
- Разработчики игр — для создания захватывающих звуковых ландшафтов и озвучивания персонажей в видеоиграх.

## Отзывы в интернете

*“Это же находка для визуальных новелл! Каждый персонаж со своим уникальным голосом был бы очень крут!”*

*“Я использую версию, установленную на сервере Discord, но Bark всё ещё очень глючный”*

*“Честно говоря, пока что это выглядит не очень хорошо. Мне нравятся идеи метатегов и переключения кода, но когда вы сравниваете даже лучший результат с материалами ElevenLabs, это совсем не тот уровень”*

## Тестовые примеры

1. Озучивание текста заданным голосом с дополнительными звуками:

- Модели было необходимо озвучить фразу на английском языке и сгенерировать дополнительные звуки. Исходный текст:  
«[clears throat] Go back, say,... 300 years. The things that we take for granted today you'd be burned at the stakeful. Being able to fly. That's crazy. [lauhgs]»

Сгенерированная речь

## 2. Озучивание текста разными голосами:

- Модели было необходимо озвучить фразу разными голосами на русском языке. Исходный текст:

MEN: Я помню чудное мгновенье: передо мной явилась ты.

WOMEN: Как мимолётное виденье, как гений чистой красоты.

### Сгенерированная речь

- Модели было необходимо озвучить фразу разными голосами на английском языке.

Исходный текст:

MEN: Hello, Kate. How are you today?

WOMEN: I'm feeling much better now. Thanks for asking.

### Сгенерированная речь

---

## 3. Генерация музыки, песни:

- Модели было необходимо сгенерировать музыку. Исходный промпт: [music]

### Сгенерированная музыка

- Модели было необходимо сгенерировать песню. Исходный промпт: ♪Happy Birthday to you. Happy Birthday to you.

### Сгенерированная песня

---

## Возможность автономного запуска

Модель Bark основана на трансформерах, как и Mistral NeMo, поэтому автономный запуск возможен. Однако, Bark использует TTS модели для генерации речи.

Исходя из этого, потребуется адаптировать архитектуру Mistral NeMo и обучающие наборы данных для достижения интересующей функциональности.

В большинстве случаев более целесообразно использовать специализированные модели для задач генерации речи, так как они уже оптимизированы для этих целей.

## Модель распространения

Полная бесплатная версия Bark представлена в виде репозитория с исходным кодом на GitHub и не ограничивает пользователя в функциональности.

Если нет возможности установить и использовать продукт локально, можно воспользоваться его демо-версией — [Ссылка на демо-версию Bark](#).

В таком случае пользователю доступна функция воспроизведения текста одним из предложенных голосов на интересующем языке. Однако выходная аудиозапись ограничена 15 секундами.

Коммерческое использование: Лицензия MIT позволяет использовать Bark в коммерческих целях без дополнительной платы.

## Преимущества Bark

- Широкий выбор языков для распознавания и клонирования, в том числе и русский;
- Есть возможность клонировать интонацию и стиль речи вместе с голосом.
- Расширенные возможности синтеза за счёт способности инструмента генерировать различные типы звуковых сигналов из текстовых подсказок и музыку;
- Suno AI Bark может быть интегрирован в существующие рабочие процессы с помощью библиотеки Hugging Face Transformers, что упрощает использование разработчиками;
- Доступна демо-версия.

## Недостатки Bark

- Модель обучена преимущественно на английской речи, из-за чего при генерации голоса на других языках слышен акцент;
- Результат работы модели бывает непредсказуем, есть погрешности в произношении, интонации, ударениях;
- Ресурсоёмкость модели — для создания высококачественного звука требуется значительный объем видеопамати;
- Голоса не всегда похожи после клонирования;
- Генерация дополнительных звуков, музыки, разделение голосов в запросе возможно только для запросов на английском языке.

## Последняя дата обновления

Последняя дата обновления на GitHub: 01.05.2023

From:  
<http://wiki.nic.etu/docuwiki/> - REC ETU Wiki

Permanent link:  
[http://wiki.nic.etu/docuwiki/doku.php/ntn:llm:ai\\_review:20241030:bark](http://wiki.nic.etu/docuwiki/doku.php/ntn:llm:ai_review:20241030:bark)

Last update: **2024/12/18 17:20**

