

ПРАКТИЧЕСКОЕ РУКОВОДСТВО



ElevenLabs

Голос вашего дела

Коханистая Виктория

Виктория К

ElevenLabs. Голос вашего дела. От первой кнопки до собственного голосового агента

<https://litres.ru/74371338>

SelfPub; 2026

Аннотация

Синтез речи перестал быть экзотикой: сегодня ролик озвучивается за четыре минуты и стоит дешевле чашки кофе. ElevenLabs умеет это — и ещё десяток вещей, о которых большинство не догадывается.

Книга ведёт от первой кнопки до собственного голосового агента. Вы научитесь озвучивать текст так, чтобы слушатель не заподозрил синтез, создадите клон своего голоса, переведёте видео на десятки языков с сохранением тембра, соберёте бота, который принимает звонки за вас, и научитесь считать кредиты, чтобы не переплачивать.

Внутри — 12 глав, 128 готовых промтов, задания и чек-листы после каждой главы, справочники по моделям, тарифам и API. Все цифры и правила сверены с официальной документацией.

Английский интерфейс не помеха, опыт программирования не нужен. Достаточно уметь пользоваться компьютером — остальное объясняется с нуля.

Содержание

Практическое руководство	6
Введение. Почему голос — это ваш следующий инструмент	7
Содержание	10
Глава 1.	14
Знакомство: что такое ElevenLabs и что он умеет	14
Глава 2	23
Старт: аккаунт, интерфейс, кредиты, тарифы и право	23
Глава 3	33
Text to Speech: первая профессиональная озвучка	33
Длинные тексты: как не потерять интонацию	39
Глава 4	42
Язык аудиотегов: промт-инжиниринг для голоса	42
Конец ознакомительного фрагмента.	43

**Виктория К
ElevenLabs. Голос
вашего дела. От первой
кнопки до собственного
голосового агента**

Практическое руководство

ИЗДАНИЕ 2026

ElevenLabs. Голос вашего дела

От первой кнопки до собственного голосового агента: полный курс по синтезу речи, клонированию голоса, дубляжу, музыке и API — для тех, кто начинает с нуля.

12 глав · 148 готовых промтов · Задания и чек-листы
· Справочник API

Все факты, цифры и ссылки сверены с официальной документацией ElevenLabs по состоянию на август 2026 года. Цены и лимиты компания периодически меняет — актуальные значения всегда на elevenlabs.io/pricing.

Введение. Почему голос — это ваш следующий инструмент

Ещё пять лет назад, чтобы озвучить пятиминутный ролик, нужно было найти диктора, согласовать ставку, дожидаться записи, отслушать, попросить переделать одну фразу — и заплатить за неё как за полную смену. Сегодня та же задача решается за четыре минуты и стоит меньше чашки кофе. Это не преувеличение и не реклама: это арифметика, которую вы сами проверите в третьей главе.

ElevenLabs — платформа, которая превращает текст в человеческую речь, речь в текст, один голос в другой, видео на русском в видео на японском, а описание «тёплый низкий женский голос, рассказчица подкаста о путешествиях» — в готовый голос, которого не существует. Внутри неё живут ещё генератор музыки, генератор звуковых эффектов, редактор аудиокниг и конструктор голосовых роботов, которые отвечают на звонки вместо вас.

ЧТО ВЫ ПОЛУЧИТЕ, ДОЧИТАВ

- Уверенно озвучите любой текст так, чтобы слушатель не заподозрил синтез.
- Создадите свой голос — клон настоящего или придуманный с нуля.
- Переведёте своё видео на 30+ языков, сохранив собственный тембр.

- Соберёте голосового ассистента, который принимает звонки и записывает клиентов.
- Научитесь считать деньги: где кредиты утекают зря, а где экономия в 2 раза.
- Заберёте библиотеку из 148 промтов, которые работают сразу, без правок.

Как устроена эта книга

Каждая глава начинается с блока **«Какие проблемы решаем»** — чтобы вы сразу видели, зачем вам эти двадцать минут. Дальше идёт разбор: что нажать, какие настройки поставить, почему именно такие. Промты вынесены в отдельные жёлтые карточки — их можно копировать целиком. В конце главы вас ждут **задания** (короткие, на 5–15 минут), **чек-лист** для самопроверки и **краткие итоги**.

Читать подряд не обязательно. Главы 1–4 — фундамент, их лучше пройти по порядку. Дальше берите то, что нужно сейчас: дубляж — глава 7, боты — глава 10, автоматизация — глава 11.

ТРИ ЧЕСТНЫХ ПРЕДУПРЕЖДЕНИЯ

Первое. Интерфейс ElevenLabs англоязычный. Русского перевода нет, и в книге все кнопки названы по-английски — так вы найдёте их с первого раза. Понимать английский не нужно: все нужные слова разобраны.

Второе. Платформа меняется быстро. Названия моделей,

цены и лимиты в книге актуальны на август 2026 года; логика работы меняется гораздо медленнее, и именно её вы усвоите.

Третье.Голос — чувствительная вещь. Клонировать чужой голос без разрешения нельзя ни юридически, ни по-человечески. Глава 12 объясняет, где проходит граница.

С чего начать прямо сейчас

Откройте elevenlabs.io в соседней вкладке и заведите бесплатный аккаунт. Он даёт 10 000 кредитов в месяц — этого хватит примерно на 10 минут озвучки, то есть на всю первую половину книги. Карта не нужна.

Содержание

Двенадцать глав и шесть приложений. Рядом с каждой главой — что именно вы научитесь делать.

Глава 1. Знакомство: что такое ElevenLabs и что он умеет

Карта всей экосистемы за один присест: четыре больших продукта, четырнадцать инструментов и то, как они соединяются между собой. Разбираем, какая задача решается каким инструментом, чтобы дальше не блуждать по меню.

Глава 2. Старт: аккаунт, интерфейс, кредиты, тарифы и право

Регистрация, первый обход интерфейса и главное — экономика платформы: что такое кредиты, сколько их съедает минута речи и почему на бесплатном тарифе нельзя монетизировать контент. Учимся считать бюджет до того, как он кончится.

Глава 3. TexttoSpeech: первая профессиональная озвучка

Выбор модели (v3, Multilingual v2, Flash v2.5), четыре ползунка настроек и то, что они реально делают со звуком. Готовим текст к озвучке: числа, даты, аббревиатуры, ссылки — типичные места, где синтез спотыкается.

Глава 4. Язык аудиотегов: промт-инжиниринг для голоса

Самая практичная глава книги. Квадратные скобки, которые заставляют голос шептать, смеяться, вздыхать и говорить с акцентом. Пунктуация как режиссура, паузы, ударения и управление произношением через IPA и словари.

Глава 5. Голоса: библиотека, дизайн, клонирование, ремикс

Десять тысяч готовых голосов и как найти среди них нужный за минуту. Дизайн голоса по текстовому описанию, мгновенное и профессиональное клонирование, требования к записи и подводные камни с акцентом.

Глава 6. Длинные форматы: диалоги, Studio, аудиокниги и подкасты

Как озвучить не абзац, а книгу. Многоголосые диалоги, монтажная линейка Studio, импорт EPUB и DOCX с автоматической разбивкой на главы, генерация подкаста из документа и экспорт в MP3, WAV и MP4.

Глава 7. Дубляж и локализация: ваше видео на десятках языков

Автоматический дубляж с сохранением вашего тембра, разделение до 32 говорящих, сила клонирования и ручная доводка в Dubbing Studio. Считаем стоимость и выбираем между быстрой и дорогой моделью.

Глава 8. Речь в текст и обработка звука: Scribe, субтитры, VoiceChanger

Расшифровка интервью и совещаний с разделением по спикерам, тайм-коды до слова, автоматические субтитры.

Плюс два инструмента-выручалочки: подмена голоса в готовой записи и очистка звука от шума.

Глава 9. Музыка и звуковые эффекты

Фоновая музыка под ролик по текстовому описанию — до пяти минут, с правами на коммерческое использование. Звуковые эффекты от шагов по гравию до брендового джингла, заикливание и структура удачного промта.

Глава 10. Голосовые агенты: боты, которые говорят и звонят

Конструктор разговорных агентов: системный промт из шести блоков, база знаний, инструменты и интеграции, подключение телефонии и виджет на сайт. Собираем администратора салона, который записывает клиентов.

Глава 11. API и автоматизация: когда мышки становятся мало

Первый запрос за пять минут — даже если вы никогда не писали код. Ключ, SDK, потоковая генерация, пакетная озвучка сотни файлов и связка с Google Таблицами, Telegram и no-code сервисами.

Глава 12. Сценарии применения, этика, безопасность и деньги

Сорок сценариев для работы, учёбы, творчества и быта — с прикидкой времени и стоимости. Где заканчивается закон и начинается дипфейк, как защитить свой голос и на чём на этой платформе реально зарабатывают.

Приложение А. Библиотека из 128 промтов

Сто двадцать восемь готовых к вставке шаблонов по одиннадцати категориям: соцсети, бизнес, обучение, аудиокниги, дизайн голоса, дубляж, музыка, звуковые эффекты, агенты и API.

Приложение Б. Словарь терминов

Пятьдесят понятий — от «аудиотега» до «диаризации» — простым языком, с указанием, в какой главе это встречается.

Приложение В. Справочник моделей

Все актуальные модели с идентификаторами, лимитами символов, задержкой и числом языков — таблица для быстрого выбора.

Приложение Г. Справочник API

Основные эндпоинты REST API с методами, назначением и примерами вызовов, плюс коды ошибок и лимиты.

Приложение Д. Кредиты, тарифы и калькулятор

Сколько стоит каждая функция, как перевести кредиты в минуты и как посчитать бюджет проекта на салфетке.

Глава 1.

Знакомство: что такое ElevenLabs и что он умеет

Прежде чем нажимать кнопки, полезно понять, где вы находитесь. Эта глава — карта территории.

КАКИЕ ПРОБЛЕМЫ РЕШАЕМ В ЭТОЙ ГЛАВЕ

«Я зашёл на сайт и растерялся: там десяток разделов, и я не понимаю, что мне нужно».

«Мне сказали, что там можно озвучивать текст. А что ещё? Вдруг я плачу за одно, а мог бы использовать пять функций».

«Чем ElevenLabs отличается от встроенного голосового движка в телефоне и от бесплатных сайтов озвучки?»

Одно предложение о том, что это такое

ElevenLabs — платформа для работы со звуком и голосом на основе искусственного интеллекта. Она умеет три базовые вещи и множество производных от них: **превращать текст в речь, превращать речь в текст и превращать один голос в другой**. Всё остальное — дубляж, аудиокни-

ги, музыка, телефонные роботы — надстройки над этими тремя операциями.

Компания основана в 2022 году Петром Домбковским и Матеушем Станишевским. Показательна история появления: оба выросли в Польше, где иностранные фильмы традиционно шли с закадровым переводом одним голосом. Желание сделать локализацию, сохраняющую голос и эмоцию актёра, выросло из этого детского раздражения. Поэтому дубляж в ElevenLabs — не побочная функция, а одна из центральных.

Чем это отличается от «робота из навигатора»

Старые синтезаторы речи склеивали заранее записанные кусочки или считали звук по формулам — отсюда механический тон. Современные модели устроены иначе: они предсказывают звучание фразы целиком, учитывая контекст. Модель «понимает», что после слова «увы» интонация падает, а перед вопросительным знаком — поднимается. Именно поэтому в ElevenLabs можно управлять не только словами, но и подачей: попросить голос вздохнуть, усмехнуться или произнести фразу шёпотом.

Практическое следствие: качество результата зависит от того, что вы пишете, а не только от того, какие ползунки двигаете. Текст для синтеза — это по сути сценарий для актёра. Отсюда и промпты, которым посвящена вся

Карта экосистемы: четыре двери

В 2026 году платформа разделена на четыре продукта. Понимание этого деления экономит массу времени: вы сразу идёте в нужную дверь.

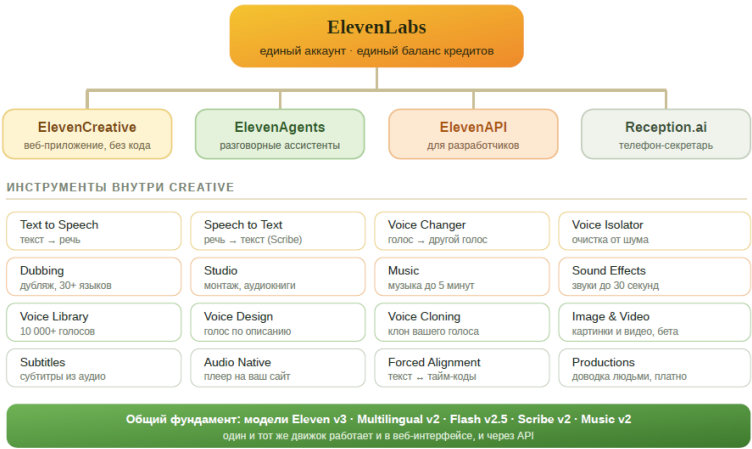


Рис. 1.1— Карта экосистемы ElevenLabs: четыре продукта, четырнадцать инструментов и общий слой моделей под ними.

ElevenCreative — там, где вы проведёте 90 % времени

Это веб-приложение, в котором ничего не нужно программировать. Открыли страницу, вставили текст, выбрали голос, нажали Generate — получили файл. Здесь же живут дубляж, монтажная линейка Studio, генератор музыки и все инструменты работы с голосами. Начинать нужно отсюда, независимо от того, кто вы по профессии.

ElevenAgents — голосовые собеседники

Отдельный конструктор, где вы описываете поведение ассистента словами, выбираете ему голос, даёте базу знаний и подключаете к телефону или сайту. Агент слушает, думает, отвечает голосом и умеет выполнять действия: записать на приём, найти заказ, перевести звонок на человека. Подробно — в главе 10.

ElevenAPI — тот же движок, но из вашей программы

REST-интерфейс с официальными библиотеками для Python и TypeScript. Нужен, когда озвучить надо не один

текст, а тысячу, или когда генерация должна происходить сама — по расписанию, по событию, внутри вашего сервиса. Глава 11 показывает, что первый рабочий запрос пишется за пять минут даже без опыта.

Reception.ai — готовый телефонный секретарь

Коробочное решение для малого бизнеса: агент, который принимает входящие звонки, отвечает на типовые вопросы и записывает клиентов. По сути это ElevenAgents с преднастроенными сценариями — берут его, когда не хочется собирать логику самому.

Что можно сделать уже сегодня: восемь честных примеров

Задача	Инструмент	Сколько времени
Озвучить сценарий Reels на 60 секунд	Text to Speech	2–4 минуты
Перевести своё видео на английский, сохранив свой голос	Dubbing	5–15 минут на ролик
Расшифровать часовое интервью с разделением по спикерам	Speech to Text (Scribe)	3–8 минут
Превратить конспект в аудиокурс для прослушивания в дороге	Studio	15–30 минут
Сделать фоновую музыку для ролика без лицензионных рисков	Music	1–3 минуты
Заменить одно неудачно произнесённое слово в готовой записи	Voice Changer	2 минуты
Посадить на сайт бота, который отвечает голосом на вопросы	ElevenAgents	40–90 минут
Озвучить 300 карточек товаров одинаковым голосом	API	20 минут на скрипт

Табл. 1.1— Типовые задачи и подходящие инструменты. Время указано для человека, прошедшего первые четыре главы.

Границы: чего платформа не делает

Честность экономит разочарования. ElevenLabs **не пишет за вас текст** — сценарий вы приносите свой (или берёте у языковой модели). Он **не монтирует видео** в полноценном смысле: Studio умеет базовую сборку, но это не Premiere. Он **не гарантирует стопроцентную идентичность** клонированного голоса — сходство высокое, но чувствительное ухо близкого человека разницу заметит. И он **не разрешает клонировать чужой голос**: профессиональное клонирование требует голосовой верификации владельца.

СОВЕТ ИЗ ПРАКТИКИ

Не пытайтесь освоить все четырнадцать инструментов сразу. Возьмите одну реальную задачу — «озвучить мой следующий ролик» — и доведите её до конца. Один завершённый результат учит больше, чем три часа кликанья по вкладкам.

ЗАДАНИЯ ДЛЯ ЗАКРЕПЛЕНИЯ

1. Откройте elevenlabs.io и зарегистрируйтесь. Ничего не генерируйте — просто найдите в левом меню разделы Text to Speech, Voice Changer, Studio, Dubbing и Music. Убедитесь, что узнаете их по рис. 1.1.

2. Выпишите на листе три свои реальные задачи, где нужен голос или звук. Напротив каждой поставьте инструмент

из табл. 1.1. Если для какой-то задачи инструмента не нашлось — отметьте её вопросом и вернитесь к ней после главы 12.

3. Определите, к какой из четырёх «дверей» вы относитесь: только веб-интерфейс (Creative), нужен бот (Agents), нужна автоматизация (API). От этого зависит, какие главы читать в первую очередь.

ЧЕК-ЛИСТ ГЛАВЫ 1

Аккаунт создан, вход в личный кабинет выполнен

Я понимаю разницу между Creative, Agents, API и Reception.ai

Я могу назвать инструмент для каждой из трёх своих задач

Я знаю, что клонирование чужого голоса без разрешения недопустимо

Я не пытаюсь освоить всё сразу и выбрал одну стартовую задачу

ЧЕМУ ВЫ НАУЧИЛИСЬ В ГЛАВЕ 1

Вы получили карту платформы и больше не теряетесь в меню. Вы знаете, что в основе всего лежат три операции — текст в речь, речь в текст, голос в голос, — а четырнадцать инструментов лишь по-разному их комбинируют. Вы понимаете, чем нейросетевой синтез отличается от старых «роботов», и почему текст для озвучки — это сценарий, а не про-

сто набор слов. И вы выбрали одну конкретную задачу, на которой будете тренироваться дальше.

Глава 2

Старт: аккаунт, интерфейс, кредиты, тарифы и право

Скучная, но самая денежная глава. Здесь вы научитесь считать — и перестанете удивляться, куда делся месячный лимит.

КАКИЕ ПРОБЛЕМЫ РЕШАЕМ В ЭТОЙ ГЛАВЕ

«Кредиты закончились на третий день, а я только пробо-вал».

«Не понимаю, какой тариф брать: разница между 6 и 99 долларами в чём?»

«Можно ли то, что я сгенерировал, выкладывать на YouTube и брать за это деньги?»

«Интерфейс на английском, я боюсь нажать не туда».

Регистрация: четыре шага

Регистрация: четыре шага

1. Откройте elevenlabs.io и нажмите **Signup** в правом верхнем углу.

2. Зарегистрируйтесь через Google или по электронной

почте. Банковская карта на бесплатном тарифе не нужна.

3. Подтвердите почту по ссылке из письма.

4. Ответьте на пару вводных вопросов о том, зачем вы пришли, — они влияют только на подсказки в интерфейсе.

Вы попадёте в личный кабинет по адресу elevenlabs.io/app. Слева — меню инструментов, сверху справа — ваш аватар и, что важнее, счётчик кредитов.

Обход интерфейса: где что лежит

Home - Стартовый экран с недавними генерациями и подборкой сценариев. Отсюда удобно возвращаться к незаконченному.

Playground - Песочница: Text to Speech, Speech to Text, Voice Changer, Sound Effects, Image & Video. Быстрые одно-разовые задачи.

Products - Крупные форматы: Studio, Dubbing, Music, Audiobooks, Transcripts, Subtitles. То, где проект живёт долго.

Voices - Ваша коллекция голосов, библиотека сообщества, Voice Design и клонирование.

Agents - Конструктор разговорных ассистентов, история звонков, аналитика.

Settings/Usage/ Billing/ APIKeys - Расход кредитов по дням, смена тарифа и создание ключа для API. Загляните сюда сразу — раздел Usage будет вашим главным индикато-

ром.

АНГЛИЙСКИЙ ИНТЕРФЕЙС — НЕ ПРОБЛЕМА

Вам понадобится примерно двадцать слов: **Generate**— создать, **Voice**— голос, **Model**— модель, **Settings**— настройки, **Download**— скачать, **Upload**— загрузить, **Credits**— кредиты, **Usage**— расход, **Stability**— стабильность, **Similarity**— сходство, **Speed**— скорость, **Addvoice**— добавить голос, **Clone**— клонировать, **Dub**— дублировать, **Project**— проект, **Chapter**— глава, **Export**— экспорт, **Preview**— предпрослушка, **Regenerate**— регенерировать, **Save**— сохранить. Всё остальное — производные.

Кредиты: единая валюта платформы

Все действия оплачиваются кредитами из общего месячного баланса. Логика простая: **один символ озвученного текста — один кредит** для флагманских моделей. Пробелы и знаки препинания считаются. Абзац в 500 знаков — это 500 кредитов, и не важно, сколько раз вы его переслушали: списание происходит в момент генерации.

Сколько кредитов съедает одна операция

за единицу измерения, указанную справа

Text to Speech — Eleven v3		1 кредит / символ
Text to Speech — Multilingual v2		1 кредит / символ
Text to Speech — Flash v2.5		0,5 кредита / символ
Voice Changer / Voice Isolator		1000 кредитов / минуту
Sound Effects (с заданной длительностью)		40 кредитов / секунду
Dubbing, Music, Speech to Text		см. табл. 2.2
Правило большого пальца: 1000 кредитов ≈ 1 минута речи ≈ 150 слов русского текста		

Рис. 2.1— *Расход кредитов по основным операциям. Flashv2.5 стоит вдвое дешевле флагманских моделей — это главный рычаг экономии.*

Переводим кредиты в минуты

Средний темп речи — около 150 слов в минуту. В русском тексте слово вместе с пробелом занимает примерно 7 знаков. Отсюда: **минута речи 1000 символов 1000 кредитов**на модели v3. Значит, бесплатные 10 000 кредитов — это около десяти минут озвучки в месяц, а тариф Creator со 121 000 кредитов — примерно два часа.

ГДЕ КРЕДИТЫ УТЕКАЮТ НЕЗАМЕТНО

Перегенерации. Каждое повторное нажатие Generate списывает полную стоимость заново. Отслушивайте на коротком фрагменте, а не на всём тексте.

Длинные предпрослушки в VoiceDesign.Текст превью тоже тарифицируется.

Дубляж сразу на пять языков.Списывается за каждый язык отдельно.

Голоса с множителем.У части голосов из библиотеки владелец когда-то установил повышенную ставку — рядом с таким голосом висит отметка о множителе кредитов.

Тарифы: что за что

Тариф	Цена в месяц	Кредитов	Что открывается
Free	\$0	10 000	Text to Speech, Speech to Text, Sound Effects, Voice Design, Music, 3 проекта. Без коммерческого использования.
Starter	\$6	30 000	Коммерческая лицензия, мгновенное клонирование голоса, 20 проектов, Dubbing Studio.
Creator	\$22	121 000	Профессиональное клонирование голоса, докупка кредитов. Первый месяц со скидкой 50 %.
Pro	\$99	600 000	Аудио 192 кбит/с, вывод PCM 44,1 кГц через API.
Scale	\$299	1 800 000	3 места в рабочем пространстве, совместная работа, 3 профессиональных клона голоса.
Business	\$990	6 000 000	10 мест, 10 профессиональных клонов, низкая задержка синтеза.
Enterprise	по договору	по договору	Индивидуальные условия, SLA, SSO, соглашения для медицинских данных, повышенные лимиты параллельных запросов.

Табл. 2.1— Тарифные планы ElevenLabs. Актуальные цифры и возможные акции — на странице тарифов.

Отдельная тарификация для API

Если вы работаете через API, компания публикует стоимость не в кредитах, а в долларах за единицу — так проще считать себестоимость сервиса.

Операция	Стоимость
Text to Speech, Eleven v3 и Multilingual v2	\$0,10 за 1000 символов
Text to Speech, Flash / v3 Conversational	\$0,05 за 1000 символов
Speech to Text, Scribe v2	\$0,22 за час аудио
Speech to Text, Scribe v2 Realtime	\$0,39 за час аудио
Дубляж v1	\$0,33 за минуту с водяным знаком, \$0,50 без него
Дубляж v2	\$2,20 за минуту
Музыка	\$0,15 за минуту
Звуковые эффекты, Voice Changer, Voice Isolator	\$0,12 за минуту
Речевой движок для агентов	\$0,08 за минуту

Табл. 2.2— Тарификация через API. Источник: elevenlabs.io/pricing/api.

Перенос и сгорание кредитов

Неизрасходованные кредиты подписки переносятся на следующий месяц, но не бесконечно: накопить можно максимум объём двух месяцев. Отдельно докупленные кредиты Pay As You Go живут 12 месяцев с момента покупки.

Какой тариф выбрать: короткое дерево решений



Рис. 2.2— *Дерево выбора тарифа. Главная развилка проходит не по объёму, а по вопросу о коммерческом использовании.*

Право: что можно публиковать

Это тот раздел, который экономит нервы и деньги.

Бесплатный тариф не даёт коммерческой лицензии. Любое использование ради заработка — реклама, платный курс, монетизированный канал — на нём запрещено.

Публикация с бесплатного тарифа требует указания источника. В названии материала должно быть упо-

минание elevenlabs.io или 11.ai. Для музыки указывается «Eleven Music».

Все платные тарифы, начиная со Starter, включают коммерческую лицензию. Сгенерированное можно использовать бессрочно, в том числе после окончания подписки, — при условии, что у вас есть права на исходный материал и вы соблюдаете правила платформы.

Бета-функции — исключение. Материалы, созданные в сервисах со статусом Beta, коммерчески использовать нельзя.

Контент, созданный вне периода платной подписки, коммерческим не становится задним числом. Сгенерировали на бесплатном, потом купили Starter — та старая генерация всё равно остаётся некоммерческой.

ПРАКТИЧЕСКИЙ ВЫВОД

Если результат пойдёт в дело — оформите Starter за \$6 до генерации, а не после. Шесть долларов дешевле, чем регенерировать весь проект.

Мини-практикум: первая генерация за 90 секунд

1. Меню слева **Playground Text to Speech**.
2. В поле голоса оставьте предложенный по умолчанию.
3. Вставьте текст: *«Проверка связи. Если вы слышите этот голос, значит, всё настроено правильно.»*
4. Нажмите **Generate** и послушайте.

5. Зайдите в **SettingsUsage** и посмотрите, сколько кредитов ушло. Запомните это число — теперь у вас есть личный масштаб.

ЗАДАНИЯ ДЛЯ ЗАКРЕПЛЕНИЯ

1. Посчитайте на калькуляторе: сколько минут речи вы получите на выбранном тарифе, если считать 1000 кредитов за минуту. Сравните с реальной потребностью — сколько минут озвучки вам нужно в месяц?

2. Сгенерируйте один и тот же абзац на модели Eleven v3 и на Flash v2.5. Сравните расход кредитов в разделе Usage и качество звучания на слух. Решите, где для вас проходит граница «экономия против качества».

3. Найдите в интерфейсе раздел Settings Usage и настройте себе привычку заглядывать туда после каждой большой генерации.

4. Ответьте письменно на один вопрос: будет ли ваш контент коммерческим? Если да — оформите платный тариф до начала работы.

ЧЕК-ЛИСТ ГЛАВЫ 2

Аккаунт подтверждён, я знаю, где счётчик кредитов

Я понимаю, что 1 символ 1 кредит, а минута речи 1000 кредитов

Я знаю, что Flash v2.5 стоит вдвое дешевле

Я нашёл раздел Settings Usage и проверил свой расход

Я определился с тарифом и понимаю, что он мне даёт

Я знаю, что на бесплатном тарифе коммерческое использование запрещено

Я знаю про требование указывать elevenlabs.io при публикации с Free

ЧЕМУ ВЫ НАУЧИЛИСЬ В ГЛАВЕ 2

Вы создали аккаунт, прошли по интерфейсу и, главное, освоили экономику платформы. Теперь вы переводите кредиты в минуты в уме, знаете четыре типичных места утечки бюджета и понимаете, за что именно доплачиваете на каждом тарифе. Вы также разобрались с юридической стороной: бесплатный тариф — только для учёбы и личных проектов, коммерция начинается со Starter. Дальше начинается собственно работа со звуком.

Глава 3

Text to Speech: первая профессиональная озвучка

Три решения определяют 90 % качества: какой голос, какая модель, какой текст. Ползунки — только четвёртое.

КАКИЕ ПРОБЛЕМЫ РЕШАЕМ В ЭТОЙ ГЛАВЕ

«Звучит как робот, хотя я ничего не менял».

«Голос читает номер телефона слитно, а сумму — как набор цифр».

«Каждый раз получается по-разному, я не могу повторить удачный дубль».

«Не понимаю, какую модель выбрать из четырёх».

Шаг первый: выбрать модель

Модель — это движок синтеза. Их несколько, и они не «лучше-хуже», а «для разного».

Модель	Для чего	Языков	Лимит символов	Задержка
Eleven v3 <code>eleven_v3</code>	Максимальная выразительность и эмоции. Понимает аудиотеги. Основной выбор для роликов, аудиокниг, рекламы.	70+	5 000	обычная
v3 Conversational <code>eleven_v3_conversational</code>	Та же выразительность, но в реальном времени — для агентов и живых диалогов.	70+	—	≈280 мс
Multilingual v2 <code>eleven_multilingual_v2</code>	Самая стабильная для длинных текстов. Лучше всех «нормализует» числа и даты. Аудиотеги не поддерживает.	29	10 000	обычная
Flash v2.5 <code>eleven_flash_v2_5</code>	Скорость и экономия: вдвое дешевле, отклик почти мгновенный. Хуже справляется со сложным текстом.	32	40 000	≈75 мс

Табл. 3.1— Модели синтеза речи и их сильные стороны. Полный перечень — в справочнике моделей.
ПРАВИЛО ВЫБОРА В ОДНУ СТРОКУ

► ПРАВИЛО ВЫБОРА В ОДНУ СТРОКУ

Ролик, реклама, книга, эмоции → v3. Длинная монотонная начитка, много цифр → Multilingual v2. Бот, живой диалог, экономия → Flash v2.5.

Шаг второй: выбрать голос

Голос влияет на результат сильнее любой настройки. Плохой голос с идеальными ползунками звучит хуже, чем хороший голос с настройками по умолчанию. Подробно про поиск и создание голосов — глава 5; пока просто откройте биб-

лиотеку и отберите три-четыре кандидата, прослушав их на одной и той же фразе из вашего реального текста.

Слушайте кандидатов на своём тексте, а не на демонстрационном. Голос, идеально читающий рекламу автомобиля, может рассыпаться на медицинской терминологии.

Шаг третий: настройки

В панели справа четыре регулятора. Рекомендованная официальной документацией отправная точка: **стабильность около 50, сходство около 75, стиль 0.**

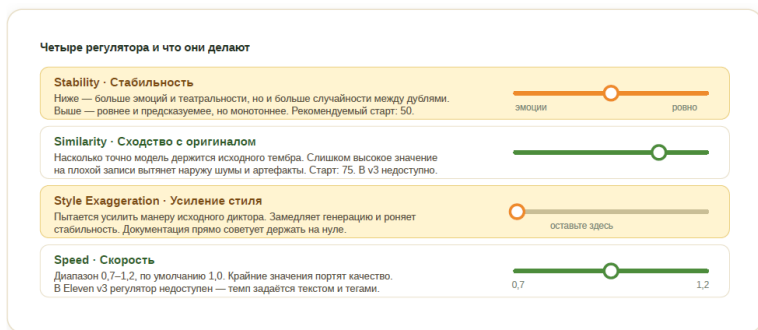


Рис. 3.1— *Регуляторы голоса в TextToSpeech. SpeakerBoost— пятый, дополнительный переключатель: слегка усиливает сходство ценой задержки, разница обычно едва слышна.*

Почему два одинаковых запроса звучат по-разному

Синтез недетерминирован. Настройки задают не точку, а **диапазон допустимой случайности** — примерно как живой актёр, читающий один и тот же текст дважды. Чем ниже стабильность, тем шире разброс. Если нужен воспроизводимый результат, работайте через API с параметром seed — но и он даёт лишь близкое, а не побитово идентичное совпадение.

Шаг четвёртый: подготовить текст

Самая недооценённая часть работы. Модель читает то, что видит, — и на сложных конструкциях спотыкается. Крупные модели справляются лучше: Multilingual v2 прочитает «\$1 000 000» как «один миллион долларов», а Flash v2.5 — как «одна тысяча тысяч долларов».

Опасные места

Что в тексте	Риск	Как писать вместо этого
+7 (999) 123-45-67	прочтёт как одно длинное число	плюс семь, девятьсот девяносто девять, сто двадцать три, сорок пять, шестьдесят семь
1 250,50 ₽	потеряет копейки или знак валюты	тысяча двести пятьдесят рублей пятьдесят копеек
25.08.2026	прочтёт по цифрам с точками	двадцать пятого августа две тысячи двадцать шестого года
ООО «ГК СтройИнвест»	прочтёт аббревиатуру по буквам невнятно	О-О-О Группа компаний Строй-Инвест
example.com/page? id=7	прочтёт спецсимволы вслух	экзампл точка ком, страница семь
9:23	прочтёт как «девять двоеточие двадцать три»	девять двадцать три

Табл. 3.2— Нормализация текста перед озвучкой: типичные ловушки и их обход.

Переписывать всё вручную не обязательно — эту работу можно поручить языковой модели. Промт ниже проверен и работает в любом чат-ассистенте.

ВАЖНО: ДАЛЕЕ В КНИГЕ ВЫ УВИДИТЕ ПРОМТЫ, КОТОРЫЕ МОЖНО ИСПОЛЬЗОВАТЬ В РАБОТЕ.

ТЕКСТ /ПРОМТА/ ВЫДЕЛЯЕТСЯ КОСЫМИ ЛИНИЯМИ.

ВСТАВЛЯТЬ В ЗАПРОС ТЕКСТ НУЖНО БЕЗ / КОСЫХ ЛИНИЙ/, ЧТОБЫ ЗАПРОС РАБОТАЛ КОР-

РЕКТНО.

##Промт 3.1 · Нормализация текста под озвучку ПОДГОТОВКА

/Преобразуй текст ниже в формат, пригодный для синтеза речи. Правила:

1. Все числа, суммы, даты, время и телефоны запиши словами так, как их произносит носитель русского языка.
2. Аббревиатуры раскрой или запиши по буквам через дефис (например: ООО О-О-О), сохранив узнаваемость.
3. Заменяй символы валют, процентов и математические знаки словами.
4. Ссылки и адреса электронной почты перепиши произносимо.
5. Не сокращай и не пересказывай текст — только переформатируй.
6. Сохрани абзацы и знаки препинания.

Верни только готовый текст, без пояснений.

ТЕКСТ:

""""

[вставьте сюда ваш текст]

""""/

Экономит 10–15 минут вычитки на каждой странице и убирает большинство ошибок произношения ещё до генерации.

Длинные тексты: как не потерять интонацию

Модель v3 принимает до 5000 символов за раз, Multilingual v2 — до 10 000. Длинный материал приходится резать. Три правила:

Режьте по смыслу, а не по счётчику. Граница куска должна совпадать с концом абзаца или сцены.

Не начинайте кусок с местоимения или союза. Модель не знает, что было до, и поставит неверную интонацию.

При работе через API передавайте соседние куски в параметрах `previous_text` и `next_text` — тогда интонация на швах не рвётся.

Форматы вывода

Доступны MP3 разного качества, WAV/PCM, μ -law и Opus. На бесплатном тарифе выдаётся MP3 128 кбит/с — этого достаточно для соцсетей. Для монтажа в видеоредакторе или для последующей обработки берите WAV (доступен с Pro). Для телефонии применяют μ -law 8 кГц — он звучит хуже, зато совместим с АТС.

ЧАСТАЯ ОШИБКА НОВИЧКА

Сгенерировать сразу весь двадцатиминутный текст, услы-

шать в третьей минуте неверное ударение и запустить всё заново. Проверяйте настройки на абзаце в 200–300 символов: это стоит 300 кредитов вместо 20 000.

Мини-практикум: озвучка ролика на 60 секунд

1. Напишите или возьмите текст на 900–1000 знаков — это примерно минута речи.
2. Прогоните его через промт 3.1.
3. Возьмите первый абзац, выберите три голоса-кандидата и сгенерируйте по одной версии каждым.
4. Выберите победителя. Поставьте стабильность 50, сходство 75, стиль 0.
5. Сгенерируйте весь текст. Прослушайте и отметьте фразы, которые звучат не так.
6. Исправьте *текст* в этих местах, а не ползунки. В девяти случаях из десяти проблема в тексте.

ЗАДАНИЯ ДЛЯ ЗАКРЕПЛЕНИЯ

1. Возьмите абзац со своей работы или учёбы, где есть числа, дата и аббревиатура. Озвучьте «как есть», послушайте. Затем прогоните через промт 3.1 и озвучьте снова. Запишите, сколько ошибок исчезло.
2. Один и тот же текст озвучьте трижды: стабильность 20, 50 и 85. Опишите словами, как менялась подача. Определите, какое значение подходит вашему жанру.

3. Сравните Eleven v3 и Multilingual v2 на тексте с большим количеством цифр. Отметьте, где различие заметно.

4. Разрежьте текст на 4000 знаков на три части по смыслу и озвучьте по очереди. Склейте в любом аудиоредакторе и проверьте, слышны ли швы.

ЧЕК-ЛИСТ ГЛАВЫ 3

Модель выбрана осознанно под задачу, а не оставлена по умолчанию

Голос отобран прослушиванием на моём реальном тексте

Стабильность 50 / сходство 75 / стиль 0 — как стартовая точка

Текст нормализован: числа, даты, аббревиатуры, ссылки

Настройки проверены на коротком фрагменте до полной генерации

Длинный текст разрезан по смыслу, а не по лимиту символов

Формат вывода соответствует назначению (MP3 для сети, WAV для монтажа)

Глава 4

Язык аудиотегов: промт-инжиниринг для голоса

Самая практичная глава книги. После

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.