

ElevenLabs or *Leo*?

The digital voice that runs on your own laptop, costs nothing per word, and sounds human.

The 250-year story of teaching machines to talk, how a free model called Leo actually works, when to pay ElevenLabs instead, and the exact kit to install Leo yourself tonight.



£0

PER WORD, FOREVER, AFTER
SETUP

9 voices

BLIND TESTED. LEO WAS
PICKED

Offline

NO ACCOUNT, NO KEY, NO
LIMITS

Why a free voice is a *big deal*

Every lesson, explainer, course and ad you make wants a voice. The best paid voices are now so good that most people cannot tell them from a person. The catch: **you pay for every single character**, and the meter never stops.

I hit that wall building The Standard, a 90-day programme with 365 written lessons. I wanted every one of them to have a "Listen" button. So I went looking for a voice that was free, private, mine to keep, and still sounded like a human being. This guide is what I found.

STEP 1

Text

You write the script, or point it at an existing lesson.

STEP 2

Leo reads

A 3-billion-parameter model performs it on your Mac's GPU.

STEP 3

Whisper checks

A second AI listens back and re-takes any slip.

STEP 4

Ship

A broadcast-loud MP3 lands in your folder.

13 months

That is how long The Standard's library (1.58 million characters, about 30 hours of audio) would take on ElevenLabs Creator: **\$22 a month for 121,000 credits, at 1 credit per character**. About \$286 of credits. With Leo it costs a bit of electricity.

This is not an anti-ElevenLabs guide. It is still the best voice money can buy, and page 9 tells you exactly when to pay for it. **The point is choice:** for the first time, a genuinely human-sounding voice is free to own.

Bellows, organs and a *singing computer*

1791

VIENNA

A machine with lungs

Wolfgang von Kempelen publishes his speaking machine: bellows for lungs, a reed for vocal cords, and a leather tube he squeezed by hand to shape vowels. It could manage short words and sounded like a haunted accordion.

THE IDEA Speech is a buzzing source shaped by a tube. Every voice in this guide still works on that principle.

1939

NEW YORK

The Voder steals the World's Fair

Homer Dudley's team at Bell Labs shows the Voder, the first fully electronic speech machine. It was not automatic: a trained operator "played" it like an organ, with keys for hiss and buzz and a foot pedal for pitch.

THE IDEA Speech can be built from electricity alone. But it took a human performer to make it sound alive.

1961

BELL LABS

A computer sings "Daisy Bell"

John Kelly and Louis Gerstman program an IBM computer to sing. Arthur C. Clarke hears the demo, and that is why HAL 9000 sings "Daisy" as he is switched off in *2001: A Space Odyssey*.

THE IDEA For the first time, a computer, not a person, decided what the voice did.

1980s

FORMANT ERA

Rules, rules, rules

Engineers hand-write the physics: "an 'ee' resonates at these frequencies", "raise the pitch at a question mark". Systems like DECtalk turn text into speech you could understand on any computer. Stephen Hawking's famous voice comes from this era.

THE LIMIT Nobody can write down every rule of how humans actually talk. So it always sounded like a robot.

Cut-and-paste humans, then the *neural leap*

2005

ATLANTA

Record a real person, chop them up

Instead of simulating speech, record it. Voice actor Susan Bennett says she spent July 2005 reading nonsense sentences for four hours a day for a speech company. Her voice was sliced into thousands of tiny sound pieces and stitched back together on demand. She later found out she was the original American voice of Siri.

THE LIMIT Great on familiar sentences, strange on anything new. A stitched sentence has no idea it is a question, a joke or a warning.

2016

DEEPMIND

WaveNet: let a neural network draw the sound wave

DeepMind builds a network that predicts audio one sample at a time, 16,000 predictions for every second of speech. In listening tests on a 1 to 5 scale it scored **4.21**, against 3.86 for the best stitched voices and 4.55 for real humans. It closed more than half the gap in one jump.

THE CATCH Gorgeous, but far too slow to run live at first. Speed has been the trade-off ever since.

2017

GOOGLE

Tacotron 2: read the text, paint the sound

Google splits the job in two. One network reads the text and paints a spectrogram, a picture of the sound's frequencies over time. A second network turns that picture into audio. Listeners struggle to tell it from a person.

WHERE YOU HAVE HEARD IT Smart speakers and sat navs. The free voice I used before Leo, Kokoro "George", is a small modern descendant of this family: clean and fast, but read aloud rather than performed.

Speech becomes a *language*

The leap that made Leo possible came from an unexpected place: the same technology as ChatGPT. Three pieces had to land first.

2022

META, GOOGLE

Sound becomes tokens

Neural audio codecs such as Meta's EnCodec (October 2022) learn to compress audio into a short stream of codes, like a vocabulary of sound. Once speech is "words", a language model can write it.

2023

MICROSOFT,
LONDON

VALL-E proves it, ElevenLabs sells it

Microsoft's VALL-E treats speech as language modelling, trained on about 60,000 hours of English, and copies a voice from a 3-second sample. Meanwhile ElevenLabs, founded in 2022 by Piotr Dąbkowski and Mati Staniszewski (reportedly inspired by badly dubbed films in their native Poland), turns this new wave into the most natural commercial voice on the market.

2024

META

Llama 3.2: a small brain anyone can download

On 25 September 2024 Meta releases Llama 3.2, including a 3-billion-parameter model small enough for a laptop, under a licence that allows free commercial use for almost everyone.

2025

CANOPY LABS

Orpheus: a language model that speaks

On 18 March 2025 Canopy Labs releases Orpheus 3B: Llama 3.2 retrained on over 100,000 hours of English speech to write audio tokens instead of words. It ships with eight voices and understands emotion tags like *sigh* and *chuckle*. One of those voices is called **Leo**.

THE MILESTONE A voice in the same league as the paid leaders, with weights you can download, keep and run offline.

How Leo turns text into a *performance*



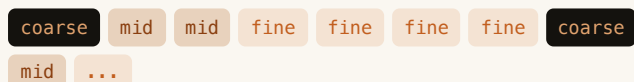
what Leo actually receives

leo: Ninety days, thirteen pillars, one small rep a day.

THE SOUND VOCABULARY

Seven tokens per slice of sound

SNAC, an open codec from researchers at Descript, describes audio at three speeds at once: one coarse code for the broad shape of the voice, two medium codes, and four fine codes for texture. Orpheus writes them in groups of seven, about **84 tokens for every second** of speech.



WHY IT SOUNDS HUMAN

Nobody programmed the inflection

There is no rule anywhere saying "rise at a question". A language model reads the whole sentence through attention before and while it speaks, so it already knows a question is coming, which word carries the stress and where a breath belongs. It learned prosody the way it learned grammar: from examples.

EMOTION ON TAP

Tags it understands

Type `<sigh>`, `<chuckle>`, `<laugh>`, `<gasp>`, `<groan>`, `<yawn>`, `<cough>` or `<sniffle>` inside your text. They work because they sat next to real sighs and laughs in the training data.

NOT A CLONE

Leo is nobody in particular

Leo is one of the eight built-in voices the model was trained to perform. It is not a copy of a real person, so there is no one's likeness to borrow and no consent to chase. The other male voices are Dan and Zac.

Why this now runs in your *laptop bag*

Ten years ago this needed a data centre. Four things changed, and all four are free or already inside your Mac.

2020 · HARDWARE

Apple Silicon's unified memory

On M-series Macs the CPU and GPU share one pool of fast memory. A multi-gigabyte model sits where the graphics chip can reach it directly, no separate graphics card needed. An M3 Pro moves data at 150 GB per second.

2023 · SOFTWARE

MLX, Apple's open AI framework

Released by Apple's machine learning research team in December 2023 under the MIT licence. The open-source mlx-audio project builds on it to run Orpheus natively on the Mac's GPU.

COMPRESSION

8-bit quantisation

Each of the model's 3.3 billion numbers is stored in 8 bits instead of 16. That halves Leo to about 3.3 GB with almost no audible loss, small enough to download once and keep.

2022 · THE PROOFREADER

OpenAI's Whisper

Released free under the MIT licence in September 2022, trained on 680,000 hours of audio. In this kit it listens to everything Leo says and flags any slip, which is how a free voice earns your trust.

3.3 GB

Leo's voice model, 8-bit

1.5 GB

Whisper, the checker

24 kHz

Output sample rate

0

Accounts, keys or logins

Offline

Once installed, the kit never touches the internet. Your scripts never leave your machine, there is no key to expire, and no company can change the price or retire the voice. **You own the whole stack.**

Why your laptop can't do it in *real time*

Leo is slower than speaking. Not because the code is bad, but because of physics. Here is the maths, so you know what to expect.

3.3 GB

Read in full for every token
Leo writes

150 GB/s

How fast an M3 Pro can
read memory

 $\approx 45/s$

Best-case tokens per
second ($3.3 \div 150 = 22$ ms
each)

84/s

Tokens Leo needs for one
second of speech

Generating speech this way is **memory-bound**, not compute-bound. To write each audio token, the chip has to stream the entire model through itself. Even in a perfect world an M3 Pro tops out at about half of real time, before the Whisper check, any re-takes, or a browser full of tabs competing for memory.

MEASURED ON	AUDIO	RENDER TIME	WHAT IT MEANS
The Standard, Day 2 lesson	5:04	18 minutes	About 3.6 times slower than real time, with every chunk checked
Same lesson, busy machine	5:04	31 minutes	Chrome holding 7 GB and swap full: close your tabs
ElevenLabs (cloud)	5:04	Typically seconds	Racks of data-centre GPUs with far more memory bandwidth

MAKE IT FASTER

- Close Chrome and other heavy apps first
- Run long jobs overnight; the kit resumes if interrupted
- Use **--no-check** for drafts to skip Whisper
- A Mac with more memory bandwidth (Max and Ultra chips) runs it proportionally faster

RULE OF THUMB

Plan on **3 to 6 minutes of rendering per minute of audio** on an M3 Pro laptop. A 20-minute course module is a coffee break. A whole library is an overnight job. Either way, it costs nothing.

ElevenLabs or Leo: the *full comparison*

	LEO (LOCAL, FREE)	ELEVENLABS (CLOUD, PAID)
MONEY		
Cost	£0 per character, forever	Starter \$6/mo for 30,000 credits. Creator \$22/mo for 121,000. 1 credit per character on its main models
One 5-min lesson	Free	About 4,200 credits: 14% of a Starter month
QUALITY		
Naturalness	Very human, real inflection and emotion tags	Best in class, most consistent over long reads
Accuracy	Occasional slips on numbers or rare words; the kit's Whisper check catches them	Rarely needs checking
Voices	8 built in (3 male, 5 female)	Thousands, plus cloning your own
Languages	English	Dozens
PRACTICALITIES		
Speed	Slower than real time on a laptop	Seconds
Setup	One script, 5 GB download, Apple Silicon Mac	Sign up and paste text
Privacy	Fully offline, text never leaves your Mac	Text is processed on their servers
Keys and limits	None. Nothing expires	API key, monthly credit cap, plan resets
Permanence	You keep the files. Nobody can retire Leo	Voices, models and prices are the vendor's call

CHOOSE LEO WHEN

- You have volume: courses, lesson libraries, archives
- You would otherwise ration credits
- Privacy matters or you want it offline
- You want a draft voice for every video before paying for a final one

PAY ELEVENLABS WHEN

- You need it fast and flawless on the first take, like an ad on a deadline
- You need another language or your own cloned voice
- You are on Windows or an older Intel Mac
- It is one short piece and your time is worth more than \$6

THE EXPERIMENT

Nine voices, one paragraph, *no names*

Before committing, I ran a blind bake-off. Every voice read the same 30-second opening of The Standard's first lesson. Every clip was matched to the same loudness (-16 LUFS) so nothing won by being louder, then shuffled and labelled Voice A to Voice I.

VOICE	TYPE	COST	WHISPER MATCH
Chris	ElevenLabs v4, the benchmark	Paid	100%
George	Kokoro, small and fast, the old default	Free	100%
Chatterbox, default	Open model, built-in voice	Free	100%
Chatterbox, expressive	Same voice, emotion turned up	Free	100%
Chatterbox, my clone	Cloned from a clip of my own voice	Free	98%
Chatterbox, my hype clip	Cloned from an energetic clip	Free	90%
Orpheus Leo	Built-in Orpheus voice	Free	96%
Orpheus Dan	Built-in Orpheus voice	Free	97%
Orpheus Zac	Built-in Orpheus voice	Free	99%

WHAT WON

Leo, by ear

Accuracy scores were close across the board, so they did not decide it. What decided it was **performance**: Leo sounded like a person explaining something he cares about, not a voice reading a page. Dan and Zac were good but slower and flatter for teaching.

THE LESSON

Test with your own script

Voices that shine on a demo sentence can tire on a 5-minute read. Run your real content, match the loudness, hide the names, and pick with your ears. The kit includes the Leo, Dan and Zac clips so you can hear them before you install anything.

The day Leo said "sixteen pounds *seventy-nine point three*"

The first full lesson I rendered was about electricity bills: a worked example full of pence, pounds and decimals. Whisper flagged five of 32 chunks. Three were harmless spelling quirks. Two were real misreads. Leo turned "£16.9973" into nonsense and repeated a whole sentence.

Paid services hide this kind of slip behind years of polish. Open models do not, so the kit adds its own polish. Here is what runs on every line you give it:

1 Short chunks

Text is split into pieces of about two sentences. Orpheus is most reliable on short inputs, and one slip never ruins a whole paragraph.

2 Numbers spelled out

"£65.80" becomes "sixty-five pounds eighty" before Leo sees it. This one change fixed both misreads.

3 Whisper listens back

Every chunk is transcribed and compared word by word with your script. Under 95% accurate, Leo reads it again with a fresh random seed, up to four takes, and the best one wins.

4 Loudness matched

Every file is normalised to -16 LUFS, the podcast standard, so it sits level next to any other audio you publish.

5 Crash-proof

Each finished chunk is saved as it completes. If the GPU hiccups, the kit restarts itself and resumes. Nothing is lost.

6 A receipt for every file

A .qa.json report lists what Whisper heard for every chunk, so you know exactly which line to listen to.

34 of 35

The re-render of that lesson with the guardrails on: 34 chunks passed first time, and the 35th was Whisper spelling "tenant's" as "tenets". **It went live on The Standard the same evening.**

INSTALL IT

Your own Leo in *one evening*

You need an Apple Silicon Mac (M1 or newer), about 6 GB of free disk and 8 GB of free memory. No coding experience: you will paste four lines into Terminal.

1 Download both zips

From the link on the next page: `leo-voice-kit.zip` (the tool, 2 MB) and `leo-voice-models.zip` (the voice, 5.3 GB).

2 Unzip them together

Unzip the kit, then unzip the models *inside* it, so you have a `leo-voice-kit/models` folder.

3 Open Terminal in the folder

Type `cd` (with a space), drag the `leo-voice-kit` folder onto the window, press Return.

4 Run setup once

Type `./setup.sh`. It installs a private Python, the exact pinned packages, and says hello in Leo's voice when it is done.

5 Make your first line

`.venv/bin/python leo.py "Hello, this is Leo."` Your MP3 appears in the `out` folder.

6 Narrate a whole script

Save your text as `script.txt` (blank line between paragraphs, # for headings) and run `.venv/bin/python leo.py script.txt -o script.mp3`.

```
# the whole thing, after unzipping
cd ~/Downloads/leo-voice-kit
./setup.sh
.venv/bin/python leo.py "Ninety days. One small rep a day. <chuckle> Let's go." -o first.mp3
```

Tonight

The one-evening challenge: install the kit, then narrate the most important page you have ever written: your About page, your best post, the first lesson of your course. Play it back to someone without telling them it is AI. Watch their face.

10 things to make with a *free voice*

01 Listen buttons on every lesson

Turn a written course into an audio course overnight.

02 Scratch tracks for ads

Cut every edit to Leo, pay for the final voice only once it is locked.

03 Audio versions of your blog

An accessibility win and a second way to be found.

04 Faceless YouTube narration

Explainers and documentaries without recording a word.

05 Your daily briefing

Have your notes, inbox summary or reading list read to you on the commute.

06 Podcast intros and outros

A consistent announcer voice for every episode.

07 Prototype phone menus and voice apps

Hear the whole flow before you build it for real.

08 Lead magnets with audio

Give your PDF a "listen instead" version.

09 Bedtime stories

Write one with your kids, then let Leo read it back.

10 Proofread by ear

Hearing your own writing aloud exposes every clumsy sentence.

USE IT WELL

Three rules

- **Never use a voice to impersonate a real person** or to deceive. Leo is a built-in voice, not a clone, so keep it that way.
- **Say it is AI** where your audience would want to know.
- **Follow Meta's Llama acceptable use policy**, which comes with the model (see the licence on the next page).

MIX AND MATCH

The smart stack

You do not have to choose one forever. The pattern that works: **Leo for volume, ElevenLabs for the flagship**. On The Standard, the very first lesson is read by Chris on ElevenLabs. The next two are read by Leo, and the rest of the library runs on free voices. Spend credits where first impressions are made.

YOUR DOWNLOAD

Get the *Leo voice kit*

Everything I use is in one shared folder: the tool, the voice, the docs and the samples, exactly as tested. Download both files and keep them. The model files are frozen at the version in this guide, so they will work the same way years from now.

FREE DOWNLOAD · GOOGLE DRIVE

Leo voice kit + models

Open the folder and download both zip files. If Google warns that the 5 GB file is too large to scan for viruses, that is normal for big files: choose "Download anyway".

Download Free

drive.google.com/drive/folders/1XLeYKJpHlcYePnec86ETcSXcgyBy_Nfv

leo-voice-kit.zip

ABOUT 2 MB · THE TOOL

leo.py (the narrator with Whisper checks, number spelling and auto-resume), setup.sh, pinned requirements, the full README, licence files, and Leo, Dan and Zac sample clips.

leo-voice-models.zip

5.3 GB · THE VOICE

Orpheus 3B in 8-bit (Leo and the other seven voices), the SNAC decoder and the Whisper checker. Unzip inside the kit folder and the kit runs fully offline.

LICENCES, IN PLAIN ENGLISH

Orpheus code (Canopy Labs): Apache 2.0. **Orpheus model weights** are built on Meta's Llama 3.2, so the **Llama 3.2 Community License** applies. It is free, including commercial use, unless your products had more than 700 million monthly users, and it comes with Meta's acceptable use policy. If you share the model yourself, include the licence and the NOTICE file from the kit, and display "Built with Llama".

Whisper (OpenAI), **SNAC** and **mlx-audio**: MIT. Full texts are in the kit's LICENSES folder. Llama 3.2 is licensed under the Llama 3.2 Community License, Copyright © Meta Platforms, Inc. All Rights Reserved.

BUILT WITH LLAMA

LAST WORD

It took 234 years to make a machine sound human. Now it *fits in your laptop.*

Von Kempelen needed bellows and two hands. Bell Labs needed a trained operator. WaveNet needed Google's data centres. You need one evening and a Mac you already own.

The paid voices are still brilliant, and you should use them where they earn their keep. But the default has changed. A voice that sounds like a person, belongs to you and costs nothing to run is now just a download away. What you build with it is the only question left.

So here is your challenge: **ship one thing with a voice this week.**

SAMUEL ILELABOYE · ELEVENLABS OR LEO? · PRICES CHECKED OCTOBER 2026