AI STACK SIGNAL

FREE PLAIN-ENGLISH GUIDE · OCT 2026

TURN YOUR COMPUTER INTO A HOME AI STUDIO

KoboldCpp is a free app that runs AI on your own computer. Its new version, 1.124, adds a song maker that turns lyrics into music, 17 new voice models (11 of them can copy a voice from a recording) and new picture models. No cloud and no subscription. This guide shows you which file to download, how to make your first song, and the fine print.

01
What it is, what's new
Songs, voices and pictures in one app.

02
The right download
Mac, Windows or Linux, and the fake site to avoid.

03
Your first song, step by step
Then a voice and a picture.

04
Prompts and fine print
Words to try, graphics cards, licenses, permission.

CHECKED AGAINST THE OFFICIAL KOBOLDCPP RELEASE NOTES, README AND WIKI

NEWS.AISTACKSIGNAL.COM/KOBOLD

01 THE SHORT VERSION WHAT IT IS

KoboldCpp is free, open-source software for running AI models on your own computer. Its README describes it as "One executable file, no installation required," with downloads for Windows, Linux and macOS. Besides chat, its README lists picture, video, speech and music features.

1 One app, one window in your browser

You start KoboldCpp, it loads an AI model, and then you use it in your web browser at `localhost:5001`. "Localhost" means the page comes from your own computer, not from a website. Music and voices have their own page at `localhost:5001/musicui`, pictures at `localhost:5001/sdui`.

2 The app and the AI models are separate

The README is clear: "Models are separate from the software." KoboldCpp is the player; the models are what it plays. The easy way to get a model is a **template**, a small `.kcppt` file. The wiki says a template downloads the model files it needs when you launch it.

3 Your stuff stays on your computer

Because the models run on your own machine, your lyrics, recordings and pictures are handled there, not sent to an AI company's server. You do need the internet to download the app and the models the first time.

Free
Open source, under the GNU AGPL v3.0 license

v1.124
Released Oct 30, 2026, 11:51 PM Central time, on GitHub

3 systems
Mac (M-series), Windows and Linux downloads

! **Only download from the real places.** The README warns that **koboldcpp.com is "a malicious fake site"**. Trust only the releases page at github.com/LostRuins/koboldcpp/releases and the official site koboldcpp.net.

01

THIS WEEK'S UPDATE

WHAT'S NEW IN 1.124

The release notes say KoboldCpp "integrated a small subset of popular models from audio.cpp," which "greatly enhances" its music and speech features. Plus new picture models.

1

Songs from your lyrics: YuE2

A music model called YuE2. You pick one of three templates:

No LLM: the lightest. "Bring your own lyrics." **Small LLM (recommended):** adds a small 4B AI that can write lyrics for you. **Big LLM:** a huge lyric-writing AI, "Only for large GPUs."

2

17 new voice models, 11 can copy a voice

Text to speech: type words, hear them spoken. The 11 marked "Voice cloning" (orange) can copy a voice from a recording:

`BreezeTTS 2` `Ming 3.0` `Qwen3-TTS Base` `VoxCPM2` `Chatterbox` `CosyVoice 3` `Fish Audio S2 Pro` `Kitten TTS 2` `OmniVoice` `PocketTTS` `VibeVoice 1.5B` `Qwen3-TTS VoiceDesign` `Chatterbox Turbo` `Maya1` `NeuTTS 2E` `Soprano 1.1 80M` `Supertonic 3`

Some are also tagged "Instructions." In the wiki's example (Qwen3-TTS VoiceDesign) that means you describe the speaker in words. MusicUI also gets a new **transcribe** tab (with Whisper models), which the notes say helps models that need a written transcript of the reference recording.

3

New picture models

MOSS Image (recommended), **Fibo 2 Edit** and **Flux 3 Klein**, each as a ready-made template. The notes also point back to Krea 2 Turbo and Qwen Image 2.1, added earlier.

4

Also in this release, for tinkerers

AI Horde now supports Chat Completions workers, a `/mcp` setup wizard and `/paste` mode for KoboldCpp Agent, plus many fixes and newly supported text models.

02 WHICH DOWNLOAD?

PICK ONE FILE

Go to `github.com/LostRuins/koboldcpp/releases/latest`, scroll to **Assets**, and pick the one file that fits your computer. That one file is the whole app.

YOUR COMPUTER, YOUR FILE

COMPUTER	DOWNLOAD	NOTES FROM THE README AND RELEASE NOTES
Mac with M1, M2, M3 or later chip	`koboldcpp-mac-arm64`	For Apple Silicon (M-series) Macs. Intel Macs have no ready-made download; the README says they need a "source build."
Windows PC with an NVIDIA graphics card	`koboldcpp.exe`	The main Windows file, made for NVIDIA users. Double-click it to open the launcher.
Windows PC without NVIDIA	`koboldcpp-nocuda.exe`	Smaller. For AMD cards the release notes recommend trying the **Vulkan** option in this build first.
Older Windows PC	`koboldcpp-oldpc.exe`	Try this if the main file won't run on an older processor or graphics card.
Linux	`koboldcpp-linux-x64`	Also comes in `-nocuda` and `-oldpc` versions, same idea as Windows.

WHAT YOU NEED

- **Internet** for the first download of the app and the models.
- **Free disk space** for the models. The sizes aren't listed in the release notes, so check before big downloads.
- **A graphics card helps.** The README says a dedicated one is optional, but the biggest models want a strong one (see the fine print).
- **A web browser** to use it.

IF YOUR MAC BLOCKS IT

- macOS may refuse to open an app it doesn't recognise.
- The README says: if macOS blocks the app, follow Apple's instructions to allow it under **System Settings**, then **Privacy & Security**.
- Apple's page: support.apple.com/en-us/102445
- Only do this for the file from the official GitHub page.

03 FROM DOWNLOAD TO A SONG
MAKE YOUR FIRST SONG

1 Download the app

Get your file from the table on page 4 and put it in your Downloads folder.

2 Download the music template

On the v1.124 release page, find "YuE2, small LLM" (marked Recommended) and save the file `YuE2-Small-LLM.kcppt` into the same folder. The wiki's tip for templates: right-click the link and save it.

3 Open KoboldCpp with the template

Windows: double-click `koboldcpp.exe` to open the launcher and load the `.kcppt` file into it (the wiki says templates can be loaded into the launcher), then click **Launch**.

Mac: open the Terminal app and type these three lines (the README's way to start it, plus the wiki's `--config` option for a template):

TYPE IN TERMINAL

```
cd ~/Downloads
chmod +x koboldcpp-mac-arm64
./koboldcpp-mac-arm64 --config YuE2-Small-LLM.kcppt
```

4 Wait while it downloads and loads

The first launch downloads the model files the template needs, so give it time and keep the window open. Keep KoboldCpp running while you use it.

5 Open the music page

In your browser go to `http://localhost:5001/musicui`. If it won't connect, the README says to wait for loading to finish.

6 Add lyrics and generate

Paste your own lyrics (or let the small AI write them), then click **Generate**. In the wiki's MusicUI walkthrough the song then appears in the right-hand tab, where you can play or download it. That walkthrough was written for an older music model, so buttons may look a little different with YuE2.

✓ **Try these lyrics** (the wiki's examples mark parts like `[Verse 1]` and `[Chorus]`): `[Verse 1] Coffee on the counter, sun is coming through / [Chorus] Saturday morning, nothing left to do`

03 NEXT EXPERIMENTS TRY A VOICE, TRY A PICTURE

To switch from music to voices or pictures, close KoboldCpp and start it again with the other model or template.

1 Try a voice

Download one voice model file (`.gguf`) from the list in the release notes, for example **Qwen3-TTS VoiceDesign** to design a voice, or **BreezeTTS 2**. In the wiki's walkthrough you pick the voice model in the launcher's **Audio** tab, launch, then open `localhost:5001/musicui` and go to its **TTS** tab. Some models may need extra files; check the model's page.

DESCRIBE THE VOICE

```
A warm, calm older woman's voice, speaking slowly like a bedtime story.
```

2 Copy a voice, only with permission

For a voice-cloning model, the wiki's walkthrough (written for Qwen3-TTS) says to set a folder of `.wav` sample recordings, "between 5 to 10 seconds long," then pick that voice in the TTS tab. Start with **your own voice**.

3 Try a picture

Save the **MOSS Image** template (marked Recommended) from the release notes and start KoboldCpp with it the same way as the music template. Then open StableUI at `localhost:5001/sdui` , type a prompt and generate. The wiki says to add a prompt and reference images for editing models; **Fibo 2 Edit** is the one with "Edit" in its name.

TYPE AS THE PICTURE PROMPT

```
A cozy home recording studio at night, warm desk lamp, guitar on a stand,
rain on the window, soft film photo look
```

IDEAS PEOPLE CAN ACTUALLY USE

- **A birthday song** with a friend's name and inside jokes in the lyrics.
- **Read your own writing aloud**: paste a blog post or a story and hear it spoken.
- **Cover pictures** for a recipe card, a newsletter or a school project.
- **Private drafts**: try ideas without uploading them anywhere.

04 BEFORE YOU RELY ON IT
THE FINE PRINT

THE CATCH

- **Big models need a strong graphics card.** The release notes call the big-LLM music template "Only for large GPUs." The README's advice: start with a smaller model if you're not sure what your computer can run.
- **Every model has its own license.** KoboldCpp itself is GNU AGPL v3.0, and bundled parts keep their own licenses. Each AI model you download has its own terms: read the model's page before using what you make in a business.
- **Only copy a voice you have permission to use.** Your own, or someone who said yes.
- **Mac may block the app** the first time. See page 4 for what the README says to do.
- **Downloads can be big** and the release notes don't list model sizes, so check your disk space.

SOURCES

1. KoboldCpp v1.124 release notes: github.com/LostRuins/koboldcpp/releases/tag/v1.124
2. KoboldCpp README (downloads, Mac note, fake-site warning, license): github.com/LostRuins/koboldcpp
3. KoboldCpp wiki (templates, MusicUI, voice cloning, StableUI): github.com/LostRuins/koboldcpp/wiki
4. Apple, opening apps safely on Mac: support.apple.com/en-us/102445

WHAT WE DON'T KNOW

- How big each model download is, and how fast songs, voices and pictures come out on an average computer.
- The exact graphics card each template needs. The notes only say the big one is for large GPUs.

news.aistacksignal.com/kobold

This guide is general information. AI Stack Signal is not connected to KoboldCpp or its makers. Facts checked against the official v1.124 release notes, README and wiki on October 31, 2026; check them for later updates.