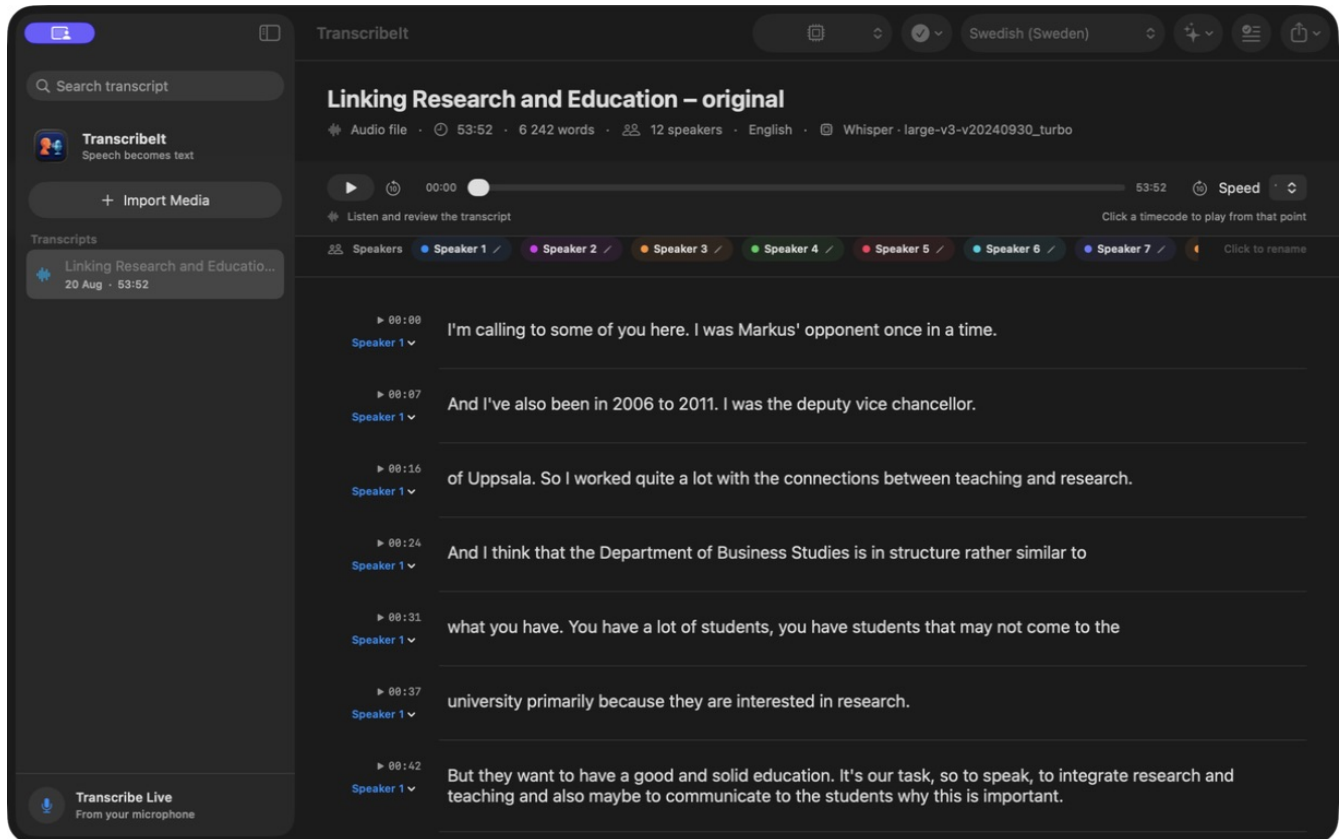


Transcribelt Local

Concise English manual - full edition 1.0

The complete workflow can run locally. After your selected models have been downloaded, transcription, speaker detection, summaries, proofreading and neural speech synthesis can be used without API credits. Internet is required for the initial model download.



Quick start

1

Install

Open the DMG and drag **Transcribelt Local** to Applications. The app requires Apple Silicon and macOS 14 or later.

2

Choose engine and language

Use the toolbar controls. **Apple Speech** is built into macOS. **Whisper** provides local transcription with a downloaded model. The system language is used on first launch.

3

Transcribe

Choose **Import Media** for audio or video, or **Transcribe Live** for the microphone. Allow microphone and speech-recognition access when macOS asks.

4

Review and export

Play the source, edit text and speaker names, create AI results, then select the required format from the sharing menu.

1. Transcribe audio, video and meetings

Import a file

- Click **Import Media** or **Choose Audio or Video**. You can also drag files into the window.
- Select one or several audio or video files. Multiple files are processed in sequence.
- The app saves a complete local working copy before analysis begins. Audio is extracted from video automatically.
- Progress appears in the app. Long files take more time, particularly with large Whisper models and speaker detection.

Live transcription

- Click **Transcribe Live** or **Start Live** and speak into the selected Mac microphone.
- Click **Stop Live** to finish. Wait for final processing - the recording is stored together with the transcript.
- When a Whisper model is selected, the final pass uses Whisper. Otherwise the app uses Apple Speech.

Engine, model and language

Choice	Best used when	Important
Apple Speech	You want a quick start without a model download.	Runs on-device when macOS supports the selected language.
Whisper	You want consistent quality across many languages.	A Whisper model must be downloaded and selected.
Automatic language	The source language is unknown or varies.	Available for Whisper in Settings - General .
Specified language	You know which language is spoken.	Usually improves spelling and reduces misinterpretation.

For the best result: use a clear microphone, reduce overlapping speech and specify the source language when known. Choose Large v3 Turbo for accuracy or Base/Small for faster tests.

Speaker detection

- Enable **Identify different speakers automatically** in **Settings - General**.
- Choose **Automatic** or a known speaker count from 1 to 10. Analysis runs after transcription.
- Speaker names can always be corrected later. Results improve when voices do not overlap and the recording is clear.

2. Review, play and edit

Synchronized playback

- Click a timecode or transcript row to jump to the matching point in the audio or video.
- Use play/pause, 10-second jumps, the timeline and speeds of 0.75x, 1x, 1.25x, 1.5x or 2x.
- For video, show or hide the movie panel. **Show Current Transcript Row** jumps from the current frame to the correct text row.
- If the media file is missing, select **Choose Media File...** to reconnect it. The app stores a new local copy.

Correct text and speakers

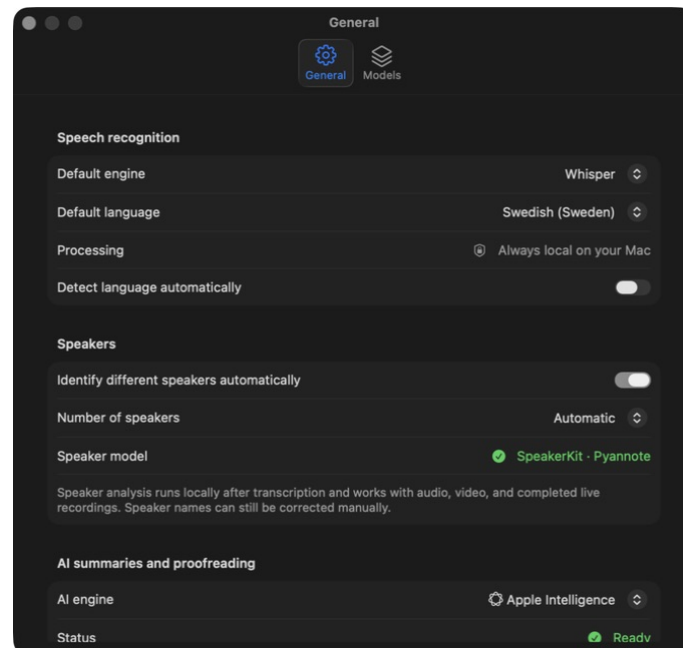
- Click the text in a row and type the correction directly.
- Click a speaker name and rename it, for example from Speaker 1 to Anna. Every segment assigned to that speaker is updated.
- Use the speaker menu on each row to select a known name or Speaker 1-8.
- Copy or delete an individual row from its menu. Incorrect or hallucinated ending segments can therefore be removed.

Organize transcripts

- Edit the title at the top to rename the transcript.
- Search your transcript collection from the sidebar search field.
- Right-click a transcript for **Copy Text** or **Delete**. Deletion cannot be undone.

Recommended workflow: listen at 0.75x or 1x, correct speaker names first, edit the text, then run AI Proofreading. Exports use the saved, corrected transcript.

General settings



General contains the default engine and language, speaker detection, AI engine and default export format.

3. AI summaries and proofreading

Select an AI engine

Engine	Benefit	Requirement
Apple Intelligence	Integrated with macOS.	A compatible Mac, enabled Apple Intelligence and language support.
Downloaded local model	Gemma runs locally without API credits.	Gemma 3 or Gemma 4 must be downloaded and selected.

Create a summary

- Open **Summarize** from the toolbar.
- Choose **Summary** and set an approximate length from 100 to 2,000 words in steps of 50. The result may vary slightly to summarize the complete source naturally.
- Choose **Meeting minutes** for topics, discussion, decisions, action items and open questions.
- Choose an output language: **Same as Transcript** or Swedish, English, German, French, Spanish, Italian, Portuguese, Dutch, Nordic languages, Polish, Czech, Ukrainian, Turkish, Arabic, Hebrew, Hindi, Japanese, Korean or Chinese.
- Long transcripts are processed in sections and combined into a complete result.
- Edit the result and use **Copy**, **Save Changes** or **Create Again**.

AI proofreading

- Open **AI Proofreading**. The AI looks for spelling and writing errors while trying to preserve content and meaning.
- The original and suggestion are displayed side by side. Select the changes you want and apply only the approved corrections.
- Always verify names, specialist terms, numbers and quotations manually.

Privacy: Gemma processes the text locally. No API key or paid credits are required. The download needs internet, but the model can be used offline afterwards.

Export AI results

- Summaries and minutes can be exported as Word (.docx), rendered PDF or a complete LaTeX document (.tex) for Overleaf.
- Save manual edits before export so the final version is used.

4. Export text and audio

Transcripts and subtitles

Format	Use
Text (.txt)	Plain, portable text.
RTF (.rtf)	Formatted text for Pages, Word and TextEdit.
Word (.docx)	Reading, editing and sharing in Word or Pages.
SRT (.srt)	Common subtitle format with timecodes.
WebVTT (.vtt)	Subtitles for the web and many video platforms.
JSON (.json)	Structured segments and metadata.
Copy All Text	Copies the complete corrected text to the clipboard.

Original audio

- Choose **Original Audio as M4A...** to export the recording or audio from imported media as M4A.
- This does not change the speech and is useful for archiving or verifying the transcript.

Corrected text as speech

- Choose **Corrected Text as Speech....** The current edited transcript is spoken and exported as M4A.
- When several people speak, each named speaker can use a different voice.

Speech engine	Best for	Settings
Apple voices	Simple, quick export in many languages.	Installed macOS voice per speaker and speech rate.
Kokoro Swedish	More natural Swedish with preset voices.	10 Swedish voices and speed 0.75-1.25.
Chatterbox V3	Voice cloning, mainly English.	Language, saved voice sample and voice per speaker. Other languages are experimental.

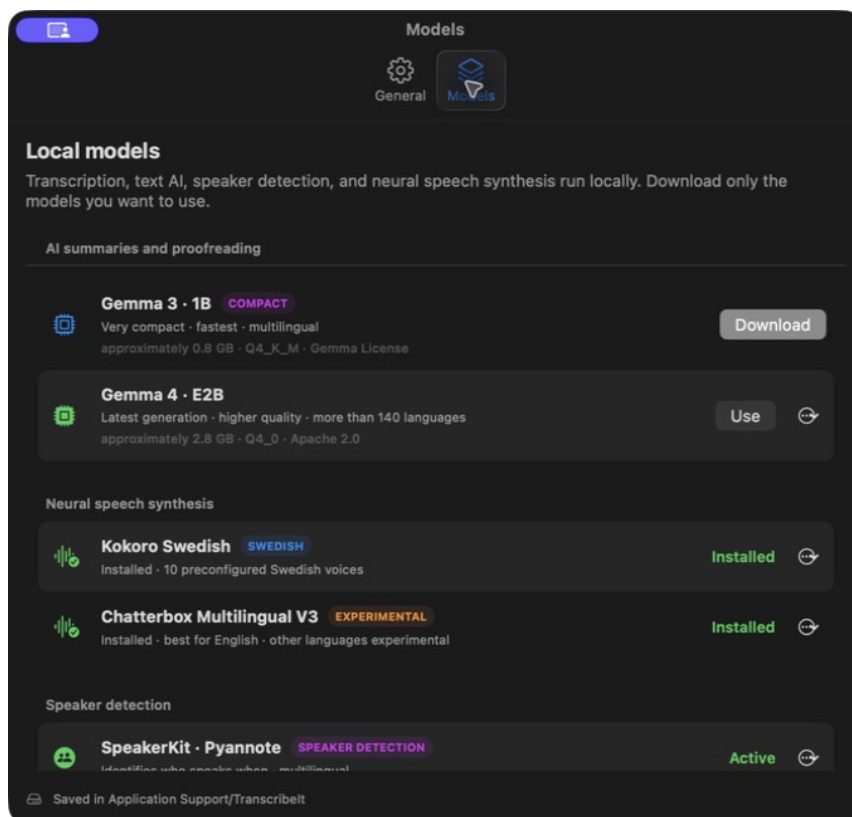
Chatterbox voice sample: use about 5-20 seconds of clear speech from one person. Save the sample in the voice library for reuse. Only use another person's voice with permission.

Quality choices

- Use Kokoro for Swedish. Chatterbox is best for English; other languages may produce an accent or silent sections.
- Preview voices before creating a long file. Longer synthesis jobs may take several minutes.

5. Settings and local models

Open **Settings** and select **Models**. Download only what you need. **Download** retrieves, **Use** selects, **Cancel** stops a download and the model menu can remove it.



Model	Size / profile	Purpose
Gemma 3 - 1B	about 0.8 GB, compact, fast	Local summaries and proofreading.
Gemma 4 - E2B	about 2.8 GB, higher quality, 140+ languages	Better local AI results, with more memory and processing time.
SpeakerKit - Pyannote	Speaker model	Identifies who speaks when.
Kokoro Swedish	Neural Swedish speech	10 preconfigured Swedish voices.
Chatterbox V3	Several GB including local engine	Voice cloning; strongest in English.

Whisper models

Model	Approximate size	Profile
Tiny / Tiny English	75 MB	Fastest; English is English-only.
Base / Base English	145 MB	Fast; useful for tests and simpler material.
Small / Small English	500 MB	Good balance between speed and quality.
Large v3 Turbo	626 MB	Recommended, multilingual and highly accurate.
Large v3 Turbo (full)	1.5 GB	Maximum quality for powerful Macs.

6. Storage, privacy and troubleshooting

Where are models and working files stored?

In Finder, choose **Go - Go to Folder...** and enter:

```
~/Library/Application Support/TranscribeIt
```

- This folder contains downloaded models, local media copies and app working data. It can grow to several GB.
- Remove models from **Settings - Models** so the app also updates their installation status.

Privacy

- Whisper, Gemma, SpeakerKit, Kokoro and Chatterbox run locally after download and use no API credits.
- Apple Speech can run on-device when the selected language is supported by macOS.
- Imported media and live recordings are stored locally for playback and export.

Common problems

Problem	Check
No live text	Microphone permission in System Settings, the selected microphone and that Stop Live was followed by final processing.
Import appears stuck	Wait for copying and final analysis. Try a smaller model. Long videos and speaker detection require more time.
Only the beginning is transcribed	Use the latest app version, allow the complete file to copy and check available storage.
Wrong language or odd words	Specify the source language or enable automatic language detection for Whisper.
Extra text after silence	Delete the incorrect row. A specified language and less background noise reduce Whisper hallucinations.
Model cannot be used	Check that the download finished, select Use and restart the app if required.
Media is missing	Select Choose Media File... and reconnect the original file.
No export dialog	Check whether the dialog is behind the window and verify access to the destination folder.

Quick recommendation: use Large v3 Turbo with automatic speaker detection for important meetings, Gemma 4 for the best local summaries and Kokoro Swedish for Swedish audio export.

Version 1.0 - manual for the full DMG edition. Availability may depend on macOS version, Mac memory and installed models.