4 Best Local AI Meeting Assistants for Mac in 2026

Compare four AI meeting assistants for Mac with documented local transcription and analysis. See what works offline, what stays on-device, and which tool fits.

Published by Julian Pscheid · June 2, 2026 · Updated August 20, 2026
Read this article online: https://www.hedy.ai/post/best-local-ai-meeting-assistants-mac/

Three professionals meeting in a small conference room, led by a Black woman speaking beside an open MacBook and a face-down phone with a subtle cyan and purple glow

Quick answer A local AI meeting assistant for Mac runs transcription and analysis on your own machine instead of a vendor's servers. All four tools here document a configuration that can do that: Hedy , Mumble, Talat, and Meetily. Hedy is the pick when you want live coaching during the conversation plus a fully local option in the same app.

Most people arrive at this question the same way. A meeting comes up where a visible bot in the participant list would be awkward, or where the transcript itself is the thing you wouldn't want sitting on someone else's servers. So you go looking for a private option, and the marketing pages all say roughly the same words: private, secure, on-device, no bots.

Those words cover at least four different architectures. A tool can skip the bot and still upload every second of audio. It can transcribe on your Mac and still send the transcript to a language model in Virginia. It can process everything locally and still sync the archive to a cloud account by default. All three are defensible designs. None of them is what "local" means to someone who wants the meeting to stay on the laptop.

This guide sorts four Mac tools by what their own documentation says about each step of the pipeline.

What does local actually mean for a meeting assistant?

Vendors bundle four separate properties into one adjective. Pull them apart before you compare anything.

Bot-free means no extra attendee joins the call. The app captures audio from your Mac's own audio system, so nothing shows up in the Zoom or Meet participant list, and the same setup works for in-person conversations. This says nothing about where the audio goes next.

Local transcription means speech-to-text runs on your hardware. It's the easiest step to move on-device and the one most often used as shorthand for the whole product. On current Apple Silicon it's close to a solved problem, which is why so many tools now advertise it.

Local analysis means the summaries, notes, and any live suggestions come from a model running on your machine. This is the expensive step, in hardware and in output quality, and it's where most tools that claim "local" quietly hand off to a cloud API.

Local storage means the recording, transcript, and notes never get copied to vendor infrastructure. It's a separate decision from processing, and it usually arrives disguised as a convenience feature called sync.

A tool that gets the first three right and syncs everything to a cloud account by default has still put your meeting in someone else's database. One that stores locally but transcribes through an API has still sent the audio. Our guide to using AI meeting notes privately (/post/how-to-use-ai-meeting-notes-privately/) walks the full data flow, including how to check the claims instead of taking them on faith.

How we picked these tools

Hedy publishes this guide, and Hedy is our product. This is an owner-published selection rather than an independent benchmark, and Hedy holds the top slot by design. Read it the way you'd read any vendor's roundup, and check the sourced facts for yourself.

We didn't install and benchmark all four. We read each vendor's current documentation, requirements pages, and security or privacy pages, then described what those primary sources say about four questions:

1. Can transcription run on the Mac?
2. Can the analysis step run on the Mac?
3. Does the tool work with the network off?
4. Where does the finished meeting get stored?

Where a vendor's claim was marketing language rather than an architectural statement, we left it out. Meetily's site, for example, describes itself as GDPR and HIPAA ready. We can't verify that, so this guide sticks to its architecture. The same standard applies to us: Hedy's entry below says what runs where and what the hardware needs are, and nothing about being the most private tool on the market.

One more disclosure, since anyone comparing local Mac tools will run into it. Mumble publishes its own roundup of local Mac meeting note takers and puts Mumble first. We're doing the same thing here. Neither list is neutral.

Prices, requirements, and platform support change often in this category, so re-check the vendor pages before you buy.

Which local AI meeting assistants work on Mac?

Tool	Fully local pipeline available	Works offline	Bot-free	Live coaching	Best for
Hedy (https://www.hedy.ai/)	Yes, opt-in on Apple Silicon	Yes, with Local AI Processing on	Yes	Yes	Real-time coaching plus a fully local option
Mumble (https://www.heymumble.com/local-mode)	Yes, in Local Mode	Yes	Yes	No	A dedicated local voice workspace for Mac
Talat (https://talat.app/)	Yes, with the built-in model or Ollama	Yes	Yes	No	Simple local transcription and a one-time-purchase path
Meetily (https://www.meetily.ai/)	Yes	Yes	Yes	No	Open-source, local-first meeting notes

Two notes on reading that table. "Live coaching" means the tool pushes suggestions to you while the conversation is still happening, and a "No" means the vendor doesn't document that feature rather than that the product is worse. "Fully local pipeline available" means a configuration exists, not that it's the default. For some of these tools, ours included, the out-of-the-box setup isn't the local one.

The four tools in detail

1. Hedy: real-time coaching with a fully local option

Hedy (https://www.hedy.ai/) is built as an AI meeting coach (/post/ai-meeting-coach/) rather than a note taker. During a session it pushes suggestions, questions worth asking, and context you might be missing, while the conversation is still moving. Transcripts, summaries, highlights, and to-dos come afterward. That live layer is why it leads this list: it's the one job on this page no other tool here documents.

The privacy architecture is a deliberate split. Speech recognition runs on-device by default in the Mac app, so audio isn't uploaded for transcription. AI analysis runs in the cloud by default, which keeps Hedy working well on hardware that can't host a language model, and cloud output is still faster and a step ahead in quality.

Then there's the switch. Local AI Processing (/help/local-ai-processing/) moves the analysis step on-device too: session summaries, detailed notes, quick prompts, chat replies, and in-session suggestions, all generated on your Mac. Speech recognition was already local, so with this turned on, the whole pipeline is. Hedy keeps working offline once the model has downloaded, and it doesn't silently fall back to the cloud. If something fails locally you get an error instead of a quiet retry against our servers. The engineering deep-dive (/post/local-ai-engineering-deep-dive-hedy-3-2/) covers how it was built and what it costs in latency.

Storage is the third decision, and it's separate. Cloud Sync is off by default. If you turn it on, Hedy copies session data to its servers so your other devices can see it. With a local speech engine, Local AI Processing on, and Cloud Sync off, nothing about the conversation reaches our servers unless you explicitly share that session.

Requirements: macOS 14.2 Sonoma or later for the app. Local AI Processing needs Apple Silicon and enough memory for the model you choose. A mid-tier model sits comfortably on a 16 GB system. The largest models need around 25 GB just to load, so 36 GB or more is realistic. Hedy flags each model as a good or poor fit for your machine, so you're not guessing.

Where it gives ground: cloud analysis is the default, and local summaries take longer than cloud ones. On a mid-tier model the wait goes from near-instant to tens of seconds or more, depending on the machine and how long the meeting ran.

Best for: anyone who wants help during high-stakes conversations, with the option to run the entire pipeline on their own hardware for the meetings that need it.

2. Mumble: a dedicated local voice workspace

Mumble (https://www.heymumble.com/local-mode) is the most single-minded local option here. Its Local Mode captures audio directly from macOS, transcribes with an on-device speech model, applies real-time speaker labels, and summarizes with a local language model. The vendor says it works with no internet connection after installation, and no bot joins the call.

It also handles dictation and voice notes, which makes it less a meeting tool than a voice workspace for the Mac. If your problem is broader than meetings, that breadth is the argument for it.

Requirements: macOS 15 or later on Apple Silicon, with 24 GB or more of memory recommended for Local Mode. That's the steepest memory recommendation in this comparison, and it's the honest cost of running capture, transcription, diarization, and summarization on one machine.

Where it gives ground: Mumble is still in beta. Its pricing page says Local Mode will remain free after beta, but the broader product and packaging can still change.

Best for: Mac users who want local voice handling across meetings, dictation, and notes, and who have the memory to spend on it.

3. Talat: simple local transcription without an account

Talat (https://talat.app/) is the least ceremonious tool on the list. It transcribes on-device as you speak, keeps recordings, transcripts, and notes in a local database that stays on the machine, identifies speakers, and needs no account to start. Every download comes with 10 hours free to try it.

For AI notes you get three paths: the built-in on-device model, a local model through Ollama, or a cloud key you supply from a provider like Claude or OpenAI. The first two keep everything on your Mac. Only the third sends anything out, and it's a choice you make rather than a behavior hidden from you.

Requirements: macOS 15 or later on Apple Silicon, with a Windows 10 or later build available too.

Pricing shape: the site currently offers monthly, annual, and one-time purchase options. The one-time path is unusual in this category, and a good fit if subscriptions are the thing you're trying to avoid.

Best for: people who want on-device transcription and a local archive without signing up for anything.

4. Meetily: open-source and local-first

Meetily (https://www.meetily.ai/) records, transcribes, and summarizes on your device, and it can run completely offline using local Whisper models for transcription and local models through Ollama for summarization. It works alongside Google Meet, Zoom, Microsoft Teams, and other conferencing apps instead of joining them, and there are macOS and Windows builds. Summaries can also use your own API key with a cloud provider, so verify which model route your setup uses.

It's open source, so the local-first claim is checkable in the code. It's also the most technical option here, and it assumes you're comfortable choosing and running your own models. The project is currently marked pre-release, and speaker diarization is listed as a Pro feature.

The site is assertive about regulatory readiness. Judge the architecture, which is genuinely local-first, and leave the compliance question to your own counsel.

Best for: technical users and small teams who want meeting notes running on infrastructure they control.

How do you choose between these?

Answer four questions in order and the list narrows fast.

How local do you actually need? All four tools document a path for running transcription and analysis on your Mac, but the local configuration isn't always the default. If the transcript itself is sensitive, check storage and sync settings too. Local processing doesn't help if the app later copies the finished session to a cloud archive.

What is your Mac? Local analysis is memory-hungry. Mumble recommends 24 GB or more for Local Mode. Hedy's mid-tier models sit comfortably on 16 GB systems, while its largest need around 25 GB just to load, making 36 GB or more realistic. Mumble and Talat require Apple Silicon on macOS 15 or later; Hedy supports macOS 14.2 or later, though Local AI Processing still requires Apple Silicon.

How much setup will you tolerate? Meetily expects you to choose and run models. Hedy, Mumble, and Talat handle model selection inside the app. Neither approach is better; they're aimed at different people.

Do you need help during the meeting or only after it? This one usually settles it. Three of these tools focus on producing a record. One also pushes suggestions while you're still in the conversation. If you're walking into a negotiation, a client call, or an interview and want a second read in the moment, that's a different product category, and it's the one Hedy is built for. Our broader AI meeting assistant comparison (/post/top-5-ai-meeting-assistants/) covers the tools that are only about the record, and our local AI productivity roundup (/post/best-local-ai-tools-for-productivity/) covers on-device tools beyond meetings.

One thing local processing doesn't solve: consent. A tool nobody can see is easier to run quietly than a bot that announces itself, which raises your obligation to say what you're running rather than lowering it. We wrote word-for-word consent scripts (/post/ask-permission-to-record-meeting-consent-scripts/) for that conversation, and the recording laws and consent (/help/recording-laws-and-consent/) article covers the jurisdictional side.

How do you set up a fully local meeting workflow on a Mac?

With Hedy it's one check and two switches.

Open Settings , then Speech & AI , then Speech Recognition Options , and confirm the speech engine is Whisper or Nemotron. Both run on your Mac, Nemotron is currently in beta, and Whisper is already the default, so most people find this one done for them.

In the same screen, turn on Local AI Processing and pick a model Hedy flags as a good fit for your machine. It's opt-in and off by default, so nothing changes until you switch it on.

Confirm Cloud Sync is off. It's off by default, but check it because turning it on copies session data to Hedy's servers for access on your other devices. With a local speech engine and Local AI Processing on, keeping Cloud Sync off closes the storage step too.

Before the meeting that actually matters, run a short session with Wi-Fi off and confirm you still get a transcript and a summary. Two minutes, and it tells you more than any vendor's privacy page, ours included.

Download Hedy for Mac (https://macos.hedy.bot/) and set up the local workflow before your next call. Local AI Processing is opt-in and off by default, so switch it on before you test Hedy offline.

Frequently asked questions

What is the best local AI meeting assistant for Mac?

It depends on which job you need done. Hedy is the pick when you want live coaching during the conversation plus a fully local option in the same app: speech recognition already runs on-device by default, and the opt-in Local AI Processing (/help/local-ai-processing/) mode moves the analysis step, meaning summaries, notes, chat, and live suggestions, onto an Apple Silicon Mac as well. Mumble is a strong dedicated local voice workspace. Talat suits people who want simple on-device transcription without an account. Meetily fits technical users and teams that want a local-first, open-source option.

Is bot-free the same as local?

No. Bot-free means no extra participant joins your call, so the app captures audio from your Mac instead of dialing in. Local means the processing happens on your hardware. A bot-free tool can still upload your audio to a transcription service and your transcript to a language model, so verify transcription, analysis, and storage separately.

Can AI meeting notes work fully offline?

Yes, on current Mac hardware. Hedy with Local AI Processing enabled, Mumble in Local Mode, Talat with its built-in model or a local one through Ollama, and Meetily with local Whisper and Ollama all document working with no connection once their models are downloaded. Offline is the useful test, because a tool that keeps producing summaries with the network off can't be sending your meeting anywhere.

Does local transcription mean summaries stay local too?

No. They're separate steps and they fail separately. Speech-to-text can run on your Mac while the transcript still goes to a cloud language model for the summary, which is what Hedy does by default. The transcript is the verbatim record, so ask where the analysis step runs and not just where the audio goes. Our guide to private AI meeting notes (/post/how-to-use-ai-meeting-notes-privately/) works through each step.

Which Mac hardware do you need for local meeting AI?

Hedy runs on macOS 14.2 Sonoma or later, and Local AI Processing needs Apple Silicon plus enough memory for the model you pick: roughly 16 GB of system memory for a mid-tier model. The largest models need around 25 GB just to load, so 36 GB or more is realistic. Mumble asks for macOS 15 or later on Apple Silicon and recommends 24 GB or more for Local Mode. Talat lists macOS 15 or later on Apple Silicon. Meetily recommends at least 16 GB of memory.

Do any of these tools store meetings only on my Mac?

Yes. Talat keeps recordings, transcripts, and notes in a local database. Meetily says all data stays on your machine when run locally, though it can also be deployed to servers you control. Mumble's Local Mode keeps the meeting on the machine. For Hedy, use a local speech engine, enable Local AI Processing, and confirm Cloud Sync is off. Cloud Sync is off by default, but disabling it alone does not stop Hedy's default cloud analysis.

How do you verify a vendor's local claim?

Start a real session with networking disabled and see whether you still get a transcript and a summary. It's the strongest single signal, though not proof, because a stall can also mean the app wants a login or a model download. Then read which processing steps the vendor names by product rather than by adjective, and check whether the app tells you when it falls back to the cloud instead of doing it quietly.

Hedy AI · Live AI Coaching for Important Conversations

Try Hedy free: https://www.hedy.ai/downloads/

https://www.hedy.ai/post/best-local-ai-meeting-assistants-mac/