

4 melhores assistentes de reuniões com IA local para Mac em 2026

Compare quatro assistentes de reuniões com IA para Mac com transcrição e análise locais documentadas. Veja o que funciona offline, o que permanece no dispositivo e qual ferramenta é a mais adequada.

Publicado por Julian Pscheid · 2 de junho de 2026 · Atualizado 20 de agosto de 2026

[Ler este artigo online: https://www.hedy.ai/pt/post/best-local-ai-meeting-assistants-mac/](https://www.hedy.ai/pt/post/best-local-ai-meeting-assistants-mac/)



Três profissionais reunidos em uma pequena sala de conferências, conduzidos por uma mulher negra que fala ao lado de um MacBook aberto e de um celular virado para baixo com um brilho sutil em ciano e roxo

Resposta rápida Um assistente de reuniões com IA local para Mac executa a transcrição e a análise na sua própria máquina, em vez de nos servidores de um fornecedor. As quatro ferramentas apresentadas aqui documentam uma configuração capaz de fazer isso: Hedy , Mumble, Talat e Meetily. Hedy é a escolha quando você quer coaching ao vivo durante a conversa e uma opção totalmente local no mesmo app.

A maioria das pessoas chega a esta pergunta pelo mesmo motivo. Surge uma reunião em que um bot visível na lista de participantes seria constrangedor, ou em que a própria transcrição é algo que você não gostaria de ver nos servidores de outra pessoa. Então você procura uma opção privada, e todas as páginas de marketing usam mais ou menos as mesmas palavras: privado, seguro, no dispositivo, sem bots.

Essas palavras abrangem pelo menos quatro arquiteturas diferentes. Uma ferramenta pode dispensar o bot e ainda enviar cada segundo de áudio. Pode transcrever no seu Mac e ainda enviar a transcrição para um modelo de linguagem na Virgínia. Pode processar tudo localmente e ainda sincronizar o arquivo

com uma conta de nuvem por padrão. As três são arquiteturas defensáveis. Nenhuma delas corresponde ao que "local" significa para quem quer manter a reunião no laptop.

Este guia organiza quatro ferramentas para Mac pelo que a documentação de cada uma diz sobre cada etapa do pipeline.

O que local realmente significa para um assistente de reuniões?

Os fornecedores agrupam quatro propriedades separadas em um único adjetivo. Separe-as antes de comparar qualquer coisa.

Sem bot significa que nenhum participante extra entra na chamada. O app captura áudio do próprio sistema de áudio do seu Mac, então nada aparece na lista de participantes do Zoom ou Meet, e a mesma configuração funciona para conversas presenciais. Isso não diz nada sobre para onde o áudio vai depois.

Transcrição local significa que a conversão de fala em texto roda no seu hardware. É a etapa mais fácil de mover para o dispositivo e a que mais costuma ser usada como atalho para o produto todo. No Apple Silicon atual, está perto de ser um problema resolvido, por isso tantas ferramentas agora anunciam isso.

Análise local significa que os resumos, notas e quaisquer sugestões ao vivo vêm de um modelo executado na sua máquina. Esta é a etapa cara, em hardware e em qualidade de saída, e é onde a maioria das ferramentas que afirmam ser "locais" transfere discretamente o trabalho para uma API de nuvem.

Armazenamento local significa que a gravação, a transcrição e as notas nunca são copiadas para a infraestrutura do fornecedor. É uma decisão separada do processamento, e geralmente chega disfarçada de recurso de conveniência chamado sincronização.

Uma ferramenta que acerta as três primeiras etapas e sincroniza tudo com uma conta de nuvem por padrão ainda colocou sua reunião no banco de dados de outra pessoa. Uma que armazena localmente, mas transcreve por uma API, ainda enviou o áudio. Nosso guia sobre como usar notas de reunião com IA de forma privada (</pt/post/how-to-use-ai-meeting-notes-privately/>) mostra todo o fluxo de dados, incluindo como verificar as alegações em vez de aceitá-las sem prova.

Como escolhemos estas ferramentas

Hedy publica este guia, e Hedy é nosso produto. Esta é uma seleção publicada pelo proprietário, não um benchmark independente, e Hedy ocupa o primeiro lugar por definição. Leia-o como leria qualquer lista de um fornecedor e verifique por conta própria os fatos com fonte.

Não instalamos nem comparamos as quatro em benchmarks. Lemos a documentação atual, as páginas de requisitos e as páginas de segurança ou privacidade de cada fornecedor, e depois descrevemos o que essas fontes primárias dizem sobre quatro perguntas:

1. A transcrição pode rodar no Mac?
2. A etapa de análise pode rodar no Mac?
3. A ferramenta funciona com a rede desligada?
4. Onde a reunião concluída é armazenada?

Quando a alegação de um fornecedor era linguagem de marketing, em vez de uma afirmação arquitetural, nós a deixamos de fora. O site da Meetily, por exemplo, se descreve como pronto para GDPR e HIPAA. Não podemos verificar isso, então este guia se limita à arquitetura dela. O mesmo

padrão se aplica a nós: a entrada da Hedy abaixo diz o que roda onde e quais são os requisitos de hardware, sem afirmar que é a ferramenta mais privada do mercado.

Mais uma divulgação, pois qualquer pessoa que compare ferramentas locais para Mac vai se deparar com isso. Mumble publica sua própria lista de tomadores de notas de reunião locais para Mac e coloca Mumble em primeiro lugar. Estamos fazendo o mesmo aqui. Nenhuma das listas é neutra.

Preços, requisitos e suporte a plataformas mudam com frequência nesta categoria, portanto verifique novamente as páginas dos fornecedores antes de comprar.

Quais assistentes de reuniões com IA local funcionam no Mac?

Ferramenta | Pipeline totalmente local disponível | Funciona offline | Sem bot | Coaching ao vivo | Ideal para

Hedy (https://www.hedy.ai/)	Sim, opcional em Apple Silicon	Sim, com o Processamento de IA Local ativado	Sim	Sim	Coaching em tempo real e uma opção totalmente local
Mumble (https://www.heyumumble.com/local-mode/)	Sim, no Local Mode	Sim	Sim	Não	Um espaço de trabalho de voz local dedicado para Mac
Talat (https://talat.app/)	Sim, com o modelo integrado ou Ollama	Sim	Sim	Não	Transcrição local simples e uma opção de compra única
Meetily (https://www.meetily.ai/)	Sim	Sim	Sim	Não	Notas de reunião open source com abordagem local-first

Duas observações para ler a tabela. "Coaching ao vivo" significa que a ferramenta envia sugestões enquanto a conversa ainda acontece, e um "Não" significa que o fornecedor não documenta esse recurso, não que o produto seja pior. "Pipeline totalmente local disponível" significa que existe uma configuração, não que ela seja a padrão. Para algumas dessas ferramentas, inclusive a nossa, a configuração pronta para uso não é a local.

As quatro ferramentas em detalhe

1. Hedy: coaching em tempo real com uma opção totalmente local

Hedy (<https://www.hedy.ai/>) foi criado como um coach de reuniões com IA (</pt/post/ai-meeting-coach/>), e não como um tomador de notas. Durante uma sessão, ele envia sugestões, perguntas que valem a pena fazer e contexto que você pode não ter, enquanto a conversa ainda está em andamento. Transcrições, resumos, destaques e tarefas vêm depois. Essa camada ao vivo é o motivo de liderar esta lista: é o único trabalho desta página que nenhuma outra ferramenta aqui documenta.

A arquitetura de privacidade é uma divisão deliberada. O reconhecimento de fala roda no dispositivo por padrão no app para Mac, portanto o áudio não é enviado para transcrição. A análise de IA roda na nuvem por padrão, o que mantém Hedy funcionando bem em hardware que não consegue hospedar um modelo de linguagem, e a saída da nuvem ainda é mais rápida e está um passo à frente em qualidade.

Depois há a opção. O Processamento de IA Local (</pt/help/local-ai-processing/>) também move a etapa de análise para o dispositivo: resumos de sessão, notas detalhadas, prompts rápidos, respostas de chat e sugestões durante a sessão, tudo gerado no seu Mac. O reconhecimento de fala já era local, portanto, com isso ativado, todo o pipeline é local. Hedy continua funcionando offline depois que o modelo é baixado e não volta silenciosamente para a nuvem. Se algo falhar localmente, você recebe um erro em vez de uma nova tentativa discreta nos nossos servidores. O mergulho técnico de engenharia (</pt/post/local-ai-engineering-deep-dive-hedy-3-2/>) explica como isso foi criado e o que custa em latência.

O armazenamento é a terceira decisão, e é separada. O Cloud Sync vem desativado por padrão. Se você o ativar, Hedy copia os dados da sessão para os servidores dela, para que seus outros dispositivos possam vê-los. Com um mecanismo de fala local, o Processamento de IA Local ativado e o Cloud Sync

desativado, nada da conversa chega aos nossos servidores, a menos que você compartilhe explicitamente essa sessão.

Requisitos: macOS 14.2 Sonoma ou posterior para o app. O Processamento de IA Local precisa de Apple Silicon e memória suficiente para o modelo escolhido. Um modelo intermediário roda com folga em um sistema de 16 GB. Os maiores modelos precisam de aproximadamente 25 GB apenas para carregar, então 36 GB ou mais é o realista. Hedy indica se cada modelo é adequado ou não para sua máquina, então você não precisa adivinhar.

Onde cede: a análise na nuvem é o padrão, e os resumos locais levam mais tempo do que os da nuvem. Com um modelo intermediário, a espera passa de quase instantânea para dezenas de segundos ou mais, dependendo da máquina e de quanto tempo a reunião durou.

Ideal para: qualquer pessoa que queira ajuda durante conversas de alto risco, com a opção de executar todo o pipeline no próprio hardware nas reuniões que precisam disso.

2. Mumble: um espaço de trabalho de voz local dedicado

Mumble (<https://www.hey-mumble.com/local-mode>) é a opção local mais focada aqui. Seu Local Mode captura áudio diretamente do macOS, transcreve com um modelo de fala no dispositivo, aplica etiquetas de interlocutores em tempo real e resume com um modelo de linguagem local. O fornecedor afirma que funciona sem conexão com a internet depois da instalação, e nenhum bot entra na chamada.

Ele também lida com ditado e notas de voz, o que o torna menos uma ferramenta de reuniões e mais um espaço de trabalho de voz para Mac. Se seu problema é mais amplo do que reuniões, essa amplitude é o argumento a favor dele.

Requisitos: macOS 15 ou posterior em Apple Silicon, com 24 GB ou mais de memória recomendados para o Local Mode. Essa é a recomendação de memória mais alta nesta comparação, e é o custo honesto de executar captura, transcrição, diarização e resumo em uma única máquina.

Onde cede: Mumble ainda está em beta. Sua página de preços diz que o Local Mode continuará gratuito depois do beta, mas o produto como um todo e o formato dos planos ainda podem mudar.

Ideal para: usuários de Mac que querem processamento de voz local em reuniões, ditado e notas, e têm memória para isso.

3. Talat: transcrição local simples sem uma conta

Talat (<https://talat.app/>) é a ferramenta menos cerimoniosa da lista. Ele transcreve no dispositivo enquanto você fala, mantém gravações, transcrições e notas em um banco de dados local que fica na máquina, identifica interlocutores e não exige conta para começar. Cada download vem com 10 horas gratuitas para experimentar.

Para notas de IA, você tem três caminhos: o modelo integrado no dispositivo, um modelo local pelo Ollama ou uma chave de nuvem que você fornece de um provedor como Claude ou OpenAI. Os dois primeiros mantêm tudo no seu Mac. Só o terceiro envia algo para fora, e é uma escolha que você faz, não um comportamento escondido de você.

Requisitos: macOS 15 ou posterior em Apple Silicon, com uma versão para Windows 10 ou posterior também disponível.

Formato de preço: o site oferece no momento opções mensal, anual e de compra única. A opção de compra única é incomum nesta categoria e adequada se assinaturas são o que você está tentando evitar.

Ideal para: pessoas que querem transcrição no dispositivo e um arquivo local sem se cadastrar em nada.

4. Meetily: open source e local-first

Meetily (<https://www.meetily.ai/>) grava, transcreve e resume no seu dispositivo, e pode rodar completamente offline usando modelos Whisper locais para transcrição e modelos locais pelo Ollama para resumos. Funciona junto com Google Meet, Zoom, Microsoft Teams e outros apps de conferência, em vez de entrar neles, e há versões para macOS e Windows. Os resumos também podem usar sua própria chave de API com um provedor de nuvem, então verifique qual caminho de modelo sua configuração usa.

É open source, portanto a alegação local-first pode ser verificada no código. Também é a opção mais técnica aqui, e pressupõe que você esteja confortável em escolher e executar seus próprios modelos. O projeto está atualmente marcado como pré-lançamento, e a diarização de interlocutores é listada como recurso Pro.

O site é assertivo sobre estar pronto para regulamentações. Avalie a arquitetura, que é de fato local-first, e deixe a questão de conformidade para sua própria assessoria jurídica.

Ideal para: usuários técnicos e pequenas equipes que querem notas de reunião rodando em uma infraestrutura que controlam.

Como escolher entre estas opções?

Responda a quatro perguntas em ordem e a lista se reduz rapidamente.

De quanto processamento local você realmente precisa? As quatro ferramentas documentam um caminho para executar transcrição e análise no seu Mac, mas a configuração local nem sempre é a padrão. Se a própria transcrição é sensível, verifique também as configurações de armazenamento e sincronização. O processamento local não ajuda se o app depois copiar a sessão concluída para um arquivo na nuvem.

Qual é o seu Mac? A análise local consome muita memória. Mumble recomenda 24 GB ou mais para o Local Mode. Os modelos intermediários do Hedy rodam com folga em sistemas de 16 GB, enquanto os maiores precisam de aproximadamente 25 GB apenas para carregar, o que torna 36 GB ou mais o realista. Mumble e Talat exigem Apple Silicon no macOS 15 ou posterior; Hedy é compatível com macOS 14.2 ou posterior, embora o Processamento de IA Local ainda exija Apple Silicon.

Quanto de configuração você tolera? Meetily espera que você escolha e execute modelos. Hedy, Mumble e Talat cuidam da seleção de modelos dentro do app. Nenhuma abordagem é melhor, pois são destinadas a pessoas diferentes.

Você precisa de ajuda durante a reunião ou apenas depois? Esta geralmente resolve a questão. Três dessas ferramentas se concentram em produzir um registro. Uma também envia sugestões enquanto você ainda está na conversa. Se você vai entrar em uma negociação, uma chamada de cliente ou uma entrevista e quer uma segunda leitura naquele momento, esta é uma categoria de produto diferente, e foi para isso que Hedy foi criado. Nossa comparação mais ampla de assistentes de reuniões com IA (</pt/post/top-5-ai-meeting-assistants/>) abrange as ferramentas que tratam apenas do registro, e nossa lista de ferramentas de IA local para produtividade (</pt/post/best-local-ai-tools-for-productivity/>) abrange ferramentas no dispositivo além de reuniões.

Uma coisa que o processamento local não resolve: consentimento. Uma ferramenta que ninguém vê é mais fácil de executar discretamente do que um bot que se anuncia, o que aumenta sua obrigação de dizer o que você está usando, em vez de diminuí-la. Escrevemos roteiros de consentimento palavra por palavra (/pt/post/ask-permission-to-record-meeting-consent-scripts/) para essa conversa, e o artigo sobre leis de gravação e consentimento (/pt/help/recording-laws-and-consent/) aborda a parte jurisdicional.

Como configurar um fluxo de trabalho de reuniões totalmente local em um Mac?

Com Hedy, é uma verificação e duas opções.

Abra Settings , depois Speech & AI , depois Speech Recognition Options , e confirme que o mecanismo de fala é Whisper ou Nemotron. Ambos rodam no seu Mac, o Nemotron está atualmente em beta e o Whisper já é o padrão, então a maioria das pessoas verá esta etapa já resolvida.

Na mesma tela, ative o Local AI Processing e escolha um modelo que Hedy indica como adequado para sua máquina. É opcional e vem desativado por padrão, então nada muda até que você o ative.

Confirme que o Cloud Sync está desativado. Ele vem desativado por padrão, mas verifique, porque ativá-lo copia os dados da sessão para os servidores da Hedy a fim de acessá-los em outros dispositivos. Com um mecanismo de fala local e o Processamento de IA Local ativado, manter o Cloud Sync desativado encerra também a etapa de armazenamento.

Antes da reunião que realmente importa, execute uma sessão curta com o Wi-Fi desligado e confirme que você ainda recebe uma transcrição e um resumo. Dois minutos, e isso diz mais do que a página de privacidade de qualquer fornecedor, inclusive a nossa.

Baixe Hedy para Mac (<https://macos.hedy.bot/>) e configure o fluxo de trabalho local antes da sua próxima chamada. O Processamento de IA Local é opcional e vem desativado por padrão, então ative-o antes de testar Hedy offline.

Perguntas frequentes

Qual é o melhor assistente de reuniões com IA local para Mac?

Depende do trabalho que você precisa fazer. Hedy é a escolha quando você quer coaching ao vivo durante a conversa e uma opção totalmente local no mesmo app: o reconhecimento de fala já roda no dispositivo por padrão, e o modo opcional Processamento de IA Local (/pt/help/local-ai-processing/) também move a etapa de análise, ou seja, resumos, notas, chat e sugestões ao vivo, para um Mac com Apple Silicon. Mumble é um ótimo espaço de trabalho de voz local dedicado. Talat é adequado para quem quer transcrição simples no dispositivo sem uma conta. Meetily serve para usuários técnicos e equipes que querem uma opção open source com abordagem local-first.

Sem bot é o mesmo que local?

Não. Sem bot significa que nenhum participante extra entra na sua chamada, então o app captura o áudio do seu Mac em vez de entrar por ligação. Local significa que o processamento acontece no seu hardware. Uma ferramenta sem bot ainda pode enviar seu áudio para um serviço de transcrição e sua transcrição para um modelo de linguagem, portanto verifique transcrição, análise e armazenamento separadamente.

Notas de reunião com IA podem funcionar totalmente offline?

Sim, no hardware atual de Mac. Hedy com o Processamento de IA Local ativado, Mumble no Local Mode, Talat com seu modelo integrado ou um modelo local pelo Ollama, e Meetily com Whisper e Ollama locais documentam que funcionam sem conexão depois que seus modelos são baixados. Offline é o teste útil, porque uma ferramenta que continua produzindo resumos com a rede desligada não pode estar enviando sua reunião para lugar nenhum.

A transcrição local significa que os resumos também permanecem locais?

Não. São etapas separadas e falham separadamente. A conversão de fala em texto pode rodar no seu Mac enquanto a transcrição ainda vai para um modelo de linguagem na nuvem gerar o resumo, que é o que Hedy faz por padrão. A transcrição é o registro literal, portanto pergunte onde a etapa de análise roda, e não apenas para onde vai o áudio. Nosso guia sobre notas de reunião privadas com IA (/pt/post/how-to-use-ai-meeting-notes-privately/) mostra cada etapa.

De qual hardware Mac você precisa para IA local em reuniões?

Hedy roda no macOS 14.2 Sonoma ou posterior, e o Processamento de IA Local exige Apple Silicon mais memória suficiente para o modelo escolhido: cerca de 16 GB de memória do sistema para um modelo intermediário. Os maiores modelos precisam de aproximadamente 25 GB apenas para carregar, então 36 GB ou mais é o realista. Mumble exige macOS 15 ou posterior em Apple Silicon e recomenda 24 GB ou mais para o Local Mode. Talat lista macOS 15 ou posterior em Apple Silicon. Meetily recomenda pelo menos 16 GB de memória.

Alguma dessas ferramentas armazena reuniões apenas no meu Mac?

Sim. Talat mantém gravações, transcrições e notas em um banco de dados local. Meetily afirma que todos os dados permanecem na sua máquina quando executado localmente, embora também possa ser implantado em servidores que você controla. O Local Mode do Mumble mantém a reunião na máquina. Com Hedy, use um mecanismo de fala local, ative o Processamento de IA Local e confirme que o Cloud Sync está desativado. O Cloud Sync vem desativado por padrão, mas desativá-lo sozinho não interrompe a análise na nuvem que Hedy faz por padrão.

Como você verifica a alegação de um fornecedor de que é local?

Inicie uma sessão real com a rede desativada e veja se você ainda recebe uma transcrição e um resumo. É o sinal isolado mais forte, embora não seja uma prova, porque uma interrupção também pode significar que o app quer um login ou o download de um modelo. Depois, leia quais etapas de processamento o fornecedor nomeia por produto, em vez de por adjetivo, e verifique se o app informa quando volta para a nuvem em vez de fazer isso discretamente.

Hedy AI · Coaching de IA ao vivo para conversas importantes

Experimente o Hedy grátis: <https://www.hedy.ai/pt/downloads/>

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