

Reconhecimento de fala no dispositivo com NVIDIA Nemotron

Nemotron é o novo motor de reconhecimento de fala no dispositivo do Hedy: totalmente local, mais preciso que motores locais anteriores e rotula quem disse o quê.

Publicado por Julian Pscheid · 23 de junho de 2026

[Ler este artigo online: https://www.hedy.ai/pt/post/nemotron-on-device-speech-engine/](https://www.hedy.ai/pt/post/nemotron-on-device-speech-engine/)



Seis colegas de diferentes etnias em uma discussão ativa ao redor de uma mesa de conferência, com uma pessoa falando e gesticulando enquanto as outras escutam

Resposta rápida Nemotron é um novo motor de reconhecimento de fala no dispositivo no Hedy, criado com base no modelo de fala Nemotron da NVIDIA. Ele roda inteiramente no seu dispositivo, transcreve com mais precisão do que motores locais anteriores e rotula quem disse o quê, de modo que uma conversa com várias pessoas fica fácil de ler depois. Você o encontrará em Settings, na seção Speech & AI, nas versões English e Multilingual. Ele se junta ao Whisper e ao Parakeet e substitui o Parakeet daqui em diante.

Dois problemas acompanham a transcrição local há muito tempo. A precisão ficava atrás da nuvem, e a transcrição de uma conversa com várias pessoas voltava como um único bloco de texto sem divisão, difícil de ler depois. Nemotron ajuda com os dois, e roda inteiramente no seu dispositivo.

Nemotron é o modelo de reconhecimento de fala no dispositivo da NVIDIA (<https://huggingface.co/nvidia/nemotron-3.5-asr-streaming-0.6b>). A NVIDIA o lançou em junho de 2026, e o Hedy está entre os primeiros aplicativos a levá-lo para um produto em produção, não apenas para uma demonstração. Agora ele fica ao lado do Whisper e do Parakeet como um dos motores de fala que você pode escolher dentro do Hedy.

Um avanço claro em precisão local

A transcrição local sempre foi a escolha certa para pessoas que preferem manter suas conversas no próprio hardware. O problema era a precisão: modelos no dispositivo ficavam atrás da nuvem, às vezes o suficiente para perceber. Nemotron reduz essa diferença. Para reuniões e chamadas do dia a dia, a transcrição que ele produz é um avanço claro em relação ao que os modelos locais conseguiam antes, sem que nada saia do seu dispositivo para isso.

Se quiser o panorama mais amplo de como a transcrição no dispositivo se encaixa no Hedy, veja nossa visão geral de IA local para reuniões (</pt/post/local-ai-meetings-hedy-3-2/>) e o mergulho técnico em engenharia do Hedy 3.2 (</pt/post/local-ai-engineering-deep-dive-hedy-3-2/>) .

Ele diferencia os seus falantes

A mudança maior está em como uma transcrição com várias pessoas aparece depois. A transcrição local mais antiga devolvia uma parede de texto sem noção de quem estava falando. Nemotron separa os falantes e os rotula: Speaker 1, Speaker 2, Speaker 3 e assim por diante. Uma reunião de vai e vem passa a ser lida como uma reunião de vai e vem.

Onde esses rótulos aparecem depende da sua plataforma:

- No iPhone e no Mac , os rótulos aparecem ao vivo enquanto as pessoas falam e a conversa alterna entre elas.
- No Windows e no Android , eles são adicionados ao final da sessão durante o processamento, então você os verá quando sua transcrição estiver pronta, em vez de em tempo real.

Dois versões: English e Multilingual

Nemotron vem em duas opções, e você escolhe a que combina com o seu trabalho. A versão English é ajustada para conversas apenas em inglês. A versão Multilingual lida com um amplo conjunto de idiomas a partir de um único modelo, e é a opção certa se suas reuniões alternam entre idiomas ou não são em inglês.

Você encontrará as duas em Settings , na seção Speech & AI , ao lado do Whisper. Trocar de motor leva apenas alguns toques. Para uma comparação lado a lado de todas as opções, veja como os motores de reconhecimento de fala do Hedy se comparam (</pt/help/speech-recognition-providers-in-hedy/>) .

Seguindo em frente depois do Parakeet

Nemotron substitui o Parakeet. Ele faz o mesmo trabalho melhor, com transcrição mais precisa. Se você usa o Parakeet hoje, mude para o Nemotron. O Parakeet está sendo retirado, e Nemotron é onde esse trabalho continua. O Whisper permanece onde está, como o motor de compatibilidade mais ampla e uma boa opção padrão se você usa um dispositivo mais antigo ou simplesmente prefere ele.

O que você precisa para rodá-lo

No iPhone e no iPad, Nemotron precisa de um iPhone 12 ou mais recente , ou de um iPad dessa geração em diante. Dispositivos Apple mais antigos não verão a opção. Ele também roda em Mac, Windows e Android.

Abra Settings !' Speech & AI , escolha Nemotron em English ou Multilingual, e sua próxima sessão será transcrita no seu dispositivo com os falantes rotulados. Mais fácil de ler depois, sem que nada disso tenha saído do seu dispositivo para acontecer.

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/nemotron-on-device-speech-engine/>