Hedy lleva NVIDIA Nemotron 3 Diarization a iPhone y Mac

Hedy 3.12 lleva Nemotron 3 Diarization de NVIDIA a iPhone y Mac con Apple Silicon, donde se ejecuta en su dispositivo para identificar quién dijo qué después de una sesión.

Publicado por Julian Pscheid · 23 de septiembre de 2026
Leer este artículo en línea: https://www.hedy.ai/es/post/nemotron-3-diarization-on-device/

Cinco colegas en una conversación animada alrededor de una amplia mesa de reuniones, con una línea luminosa de sonido por cada voz que cruza la mesa hasta un MacBook Pro

Respuesta rápida Nemotron 3 Diarization de NVIDIA es un modelo que determina quién habló y cuándo en una conversación. Llega a los iPhone y Mac con Apple Silicon compatibles con Hedy 3.12, donde se ejecuta en su dispositivo después de una sesión y actualiza qué hablante dijo cada cosa. Las palabras y las marcas de tiempo de su transcripción permanecerán iguales, y las etiquetas de hablante en vivo seguirán funcionando como hasta ahora.

NVIDIA lanzó hoy Nemotron 3 Diarization, su modelo de diarización de hablantes más avanzado hasta ahora, y Hedy lo ofrece desde el día 0 en Hedy 3.12 para iPhone y Mac. Es el modelo de diarización más preciso que hemos visto y se ejecuta en su propio dispositivo, por lo que es más fácil confiar en el registro de quién dijo qué después de una reunión.

Las etiquetas de hablante convierten una transcripción en una conversación. Cuando se desajustan, una cita se atribuye a la persona equivocada y el resumen repite el error. Ese es el problema que este modelo busca resolver.

¿Qué hace Nemotron 3 Diarization?

La diarización responde una pregunta sobre una grabación: quién habló y cuándo. Separa las voces en el audio y marca cuándo habla cada persona. Transcribir las palabras es una tarea distinta, igual que saber el nombre de cada persona.

Nemotron 3 Diarization es el modelo más reciente de NVIDIA para esa tarea. Se basa en la investigación de NVIDIA sobre Streaming Sortformer (https://arxiv.org/abs/2507.18446) , que mantiene estable la etiqueta de cada hablante a medida que avanza una conversación. Según la documentación del modelo de NVIDIA:

• Se entrenó con audio de conversaciones, incluidas reuniones, llamadas telefónicas y podcasts.

• Distingue el habla superpuesta, esos momentos en que dos personas hablan a la vez.

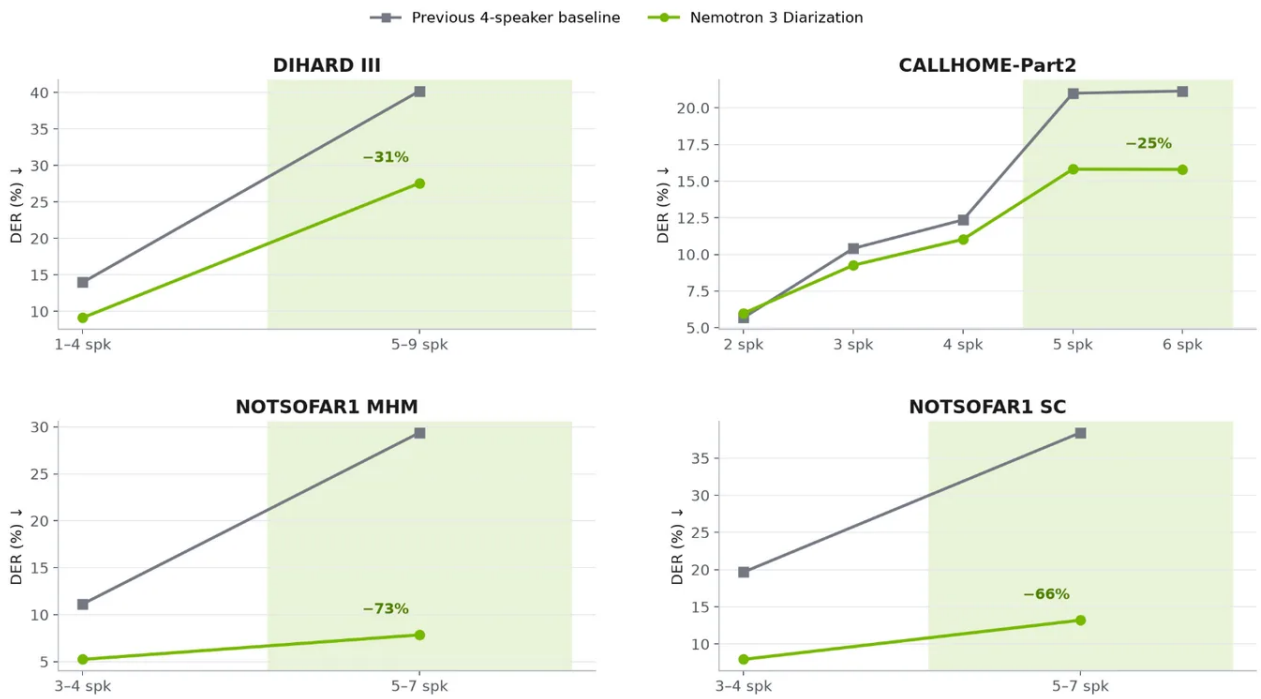
• Puede seguir hasta a ocho hablantes en una grabación.

• Etiqueta a los hablantes en el orden en que hablan por primera vez, así que la primera voz se convierte en Speaker 1.

• Es un modelo compacto, de unos 100 millones de parámetros, lo que hace viable ejecutarlo en un teléfono.

También supera a los servicios en la nube en dos pruebas de referencia publicadas. Según la publicación de lanzamiento de NVIDIA (https://huggingface.co/blog/nvidia/nemotron-diarization) , el modelo ocupa el primer puesto en la clasificación de diarización de VoiceArena (https://voicearena.com/diarization-bench) con una tasa de error de diarización del 14,72 % (la proporción de tiempo de habla etiquetado incorrectamente), por delante del siguiente sistema, con un 19,3 %. En la prueba de referencia publicada por Argmax (https://github.com/argmaxinc/OpenBench/blob/update-diarization-benchmarks/BENCHMARKS.md#diarization-error-rate-der) , obtuvo la tasa de error más baja de los seis sistemas evaluados, tres de ellos servicios de diarización en la nube, mientras se ejecutaba en un Mac.

NVIDIA también informa de mejoras frente a su modelo de diarización anterior en todas las pruebas de referencia que realizó, y la diferencia aumenta cuando hablan más de cuatro personas, situación en la que las etiquetas de hablante solían fallar.



Gráficos de líneas de NVIDIA que comparan la tasa de error de diarización según el número de hablantes en cuatro pruebas de referencia. Nemotron 3 Diarization se mantiene muy por debajo del modelo anterior de NVIDIA para cuatro hablantes, y la diferencia es mayor con cinco hablantes o más.

Gráfico: NVIDIA, de la publicación de lanzamiento de Nemotron 3 Diarization (https://huggingface.co/blog/nvidia/nemotron-diarization) . Cuanto más baja sea la tasa, mejor; las zonas sombreadas corresponden a más de cuatro hablantes. NVIDIA publica el modelo en Hugging Face (https://huggingface.co/nvidia/Nemotron-3-Diarization) junto con todos los resultados de su evaluación.

¿Qué cambia en Hedy con Nemotron 3 Diarization?

Hedy ya etiqueta a los hablantes mientras conversan mediante el motor de voz Nemotron (/post/nemotron-on-device-speech-engine/) . Esas etiquetas en vivo permanecerán exactamente como están.

La novedad ocurre después de que usted termina una sesión. Las etiquetas en vivo deben decidirse en el momento, así que Hedy ahora vuelve a examinar la conversación una vez terminada. Analiza la sesión con Nemotron 3 Diarization y actualiza a qué hablante corresponde cada parte de la transcripción.

Ese análisis solo modifica la atribución de los hablantes. Todas las palabras de su transcripción y todas las marcas de tiempo permanecen iguales. Cuando termine, los chips de hablante encima de su transcripción (/help/understanding-hedy-transcripts/) funcionarán como ahora: toque uno para darle un nombre a ese hablante o fusione dos chips si resultan ser la misma persona.

¿Se ejecuta Nemotron 3 Diarization en el dispositivo?

El objetivo de Hedy es ayudarle durante las reuniones escuchando en tiempo real y manteniendo privada su conversación. Por eso el reconocimiento de voz se ejecuta en su dispositivo de forma predeterminada, al igual que el etiquetado de hablantes.

Nemotron 3 Diarization sigue la misma regla, y nos entusiasma que un modelo tan capaz sea lo bastante pequeño para ejecutarse en un teléfono. Se ejecuta en su iPhone o Mac mediante Core ML de Apple, y el audio no se envía a ningún servicio en la nube para determinar quién estaba hablando. Para

realizar el análisis después de una sesión, Hedy conserva el audio de la sesión en una caché temporal en el dispositivo y lo elimina después. La guía de reconocimiento de voz (/help/speech-recognition-providers-in-hedy/) explica esa caché y el ajuste que la controla.

El paso de IA que redacta su resumen y sus notas es una elección aparte. También puede mantenerlo en su dispositivo con el Procesamiento de IA Local (/help/local-ai-processing/) .

¿Qué dispositivos son compatibles con Nemotron 3 Diarization?

Nemotron 3 Diarization se ejecuta en iPhone y Mac con Apple Silicon compatibles que tengan Hedy 3.12 o una versión posterior. Hedy en Windows y Android conserva sus etiquetas de hablante actuales.

Hedy 3.12 se está distribuyendo durante las próximas 24 horas, así que busque la actualización de Hedy. Cuando la tenga, abra Configuración !’ Voz e IA e instale la actualización del modelo de voz que Hedy le ofrece allí.

Se aplica a las sesiones capturadas con el motor de voz Nemotron, el que separa a los hablantes en Hedy. Si actualmente usa Whisper o un proveedor de reconocimiento de voz en la nube, cambiar a Nemotron en esa misma pantalla de configuración le dará etiquetas de hablante tanto en vivo como después de la sesión.

Preguntas frecuentes

¿Qué es Nemotron 3 Diarization?

Es un modelo de diarización de hablantes de NVIDIA. La diarización determina quién habló y cuándo en una grabación. Nemotron 3 Diarization está diseñado para conversaciones como reuniones y llamadas, distingue a las personas que hablan al mismo tiempo y puede seguir hasta a ocho hablantes.

¿Qué precisión tiene Nemotron 3 Diarization?

Según NVIDIA, ocupa el primer puesto en la clasificación de diarización de VoiceArena con una tasa de error de diarización del 14,72 %, por delante del siguiente sistema, con un 19,3 %. En la prueba de referencia publicada por Argmax, obtuvo la tasa de error más baja de los seis sistemas evaluados, incluidos tres servicios de diarización en la nube. La tasa de error de diarización es la proporción de tiempo de habla atribuido al hablante equivocado, omitido o marcado erróneamente como habla, así que cuanto más baja sea, mejor.

¿Cambiarán las etiquetas de hablante durante una sesión en vivo?

No. Las etiquetas que ve en vivo durante una sesión seguirán funcionando como hasta ahora. Nemotron 3 Diarization actúa después de que termina la sesión.

¿Se ejecuta en mi dispositivo?

Sí. El modelo se ejecuta en su iPhone o Mac mediante Core ML de Apple. Su audio no se envía a un servicio en la nube para determinar quién estaba hablando.

¿A qué sesiones se aplica?

A las sesiones capturadas con el motor de voz Nemotron. El modelo lee el audio de la sesión, así que necesita que esté en su dispositivo: debe estar activado el guardado de audio o la caché temporal de audio de Nemotron. La caché está activada de forma predeterminada y se elimina después del procesamiento.

¿Sabe cómo se llaman las personas?

No. El modelo distingue las voces y las etiqueta como Speaker 1, Speaker 2, etc. Los nombres siguen viniendo de Hedy, que los obtiene de las presentaciones, la invitación de su calendario y el contexto del tema, además de los nombres que usted escriba.

¿Qué dispositivos son compatibles?

Los iPhone y Mac con Apple Silicon compatibles que ejecutan Hedy 3.12 o una versión posterior. Hedy en Windows y Android conserva sus etiquetas de hablante actuales.
