

Hedy apporte NVIDIA Nemotron 3 Diarization sur iPhone et Mac

Hedy 3.12 intègre Nemotron 3 Diarization de NVIDIA sur iPhone et Mac avec Apple Silicon. Il s'exécute sur votre appareil pour déterminer qui a dit quoi après une session.

Publié par Julian Pscheid · 23 septembre 2026

[Lire cet article en ligne: https://www.hedy.ai/fr/post/nemotron-3-diarization-on-device/](https://www.hedy.ai/fr/post/nemotron-3-diarization-on-device/)



Cinq collègues en pleine conversation autour d'une grande table de réunion, avec une ligne sonore lumineuse distincte pour chaque voix qui traverse la table jusqu'à un MacBook Pro

Réponse rapide Nemotron 3 Diarization de NVIDIA est un modèle qui détermine qui a parlé et à quel moment dans une conversation. Il arrive avec Hedy 3.12 sur les iPhone et les Mac avec Apple Silicon compatibles, où il s'exécute sur votre appareil après la fin d'une session et met à jour l'intervenant auquel chaque passage est attribué. Les mots et les horodatages de votre transcription resteront inchangés, et les étiquettes d'intervenants en direct continueront de fonctionner comme aujourd'hui.

NVIDIA a publié aujourd'hui Nemotron 3 Diarization, son modèle de diarisation des intervenants le plus performant à ce jour, et Hedy le prend en charge dès le jour 0 dans Hedy 3.12 sur iPhone et Mac. C'est le modèle de diarisation le plus précis que nous ayons vu, et il s'exécute sur votre propre appareil, ce qui rend plus fiable l'attribution des propos après une réunion.

Les étiquettes d'intervenants transforment une transcription en conversation. Si elles dérivent, une citation est attribuée à la mauvaise personne et le résumé reprend cette erreur. C'est le problème que ce modèle cherche à résoudre.

Que fait Nemotron 3 Diarization ?

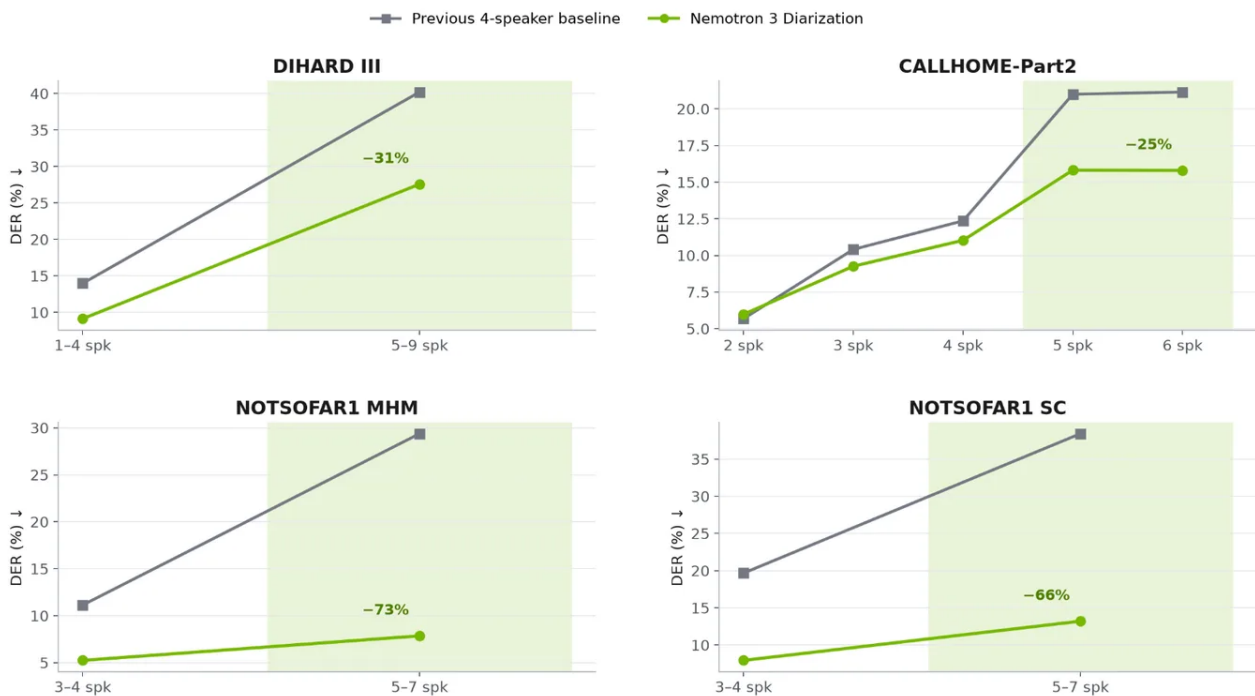
La diarisation répond à une question concernant un enregistrement : qui a parlé et à quel moment ? Elle distingue les intervenants dans l'audio et indique quand chacun parle. Transcrire les mots est une tâche distincte, tout comme connaître le nom des personnes.

Nemotron 3 Diarization est le modèle le plus récent de NVIDIA pour cette tâche. Il s'appuie sur les recherches de NVIDIA sur Streaming Sortformer (<https://arxiv.org/abs/2507.18446>) , qui maintiennent une étiquette stable pour chaque intervenant au fil de la conversation. Selon la documentation du modèle publiée par NVIDIA :

- Il a été entraîné sur des conversations audio, notamment des réunions, des appels téléphoniques et des podcasts.
- Il gère les chevauchements de parole, lorsque deux personnes parlent en même temps.
- Il peut suivre jusqu'à huit intervenants dans un enregistrement.
- Il étiquette les intervenants dans l'ordre de leur première prise de parole : la première voix devient Speaker 1.
- C'est un modèle compact, d'environ 100 millions de paramètres, ce qui rend réaliste son exécution sur un téléphone.

Il surpasse aussi les services cloud dans deux benchmarks publiés. Selon l'annonce de NVIDIA (<https://huggingface.co/blog/nvidia/nemotron-diarization>) , le modèle occupe la première place du classement de diarisation de VoiceArena (<https://voicearena.com/diarization-bench>) avec un taux d'erreur de diarisation de 14,72 % (la part du temps de parole attribuée au mauvais intervenant), devant le système suivant à 19,3 %. Dans le benchmark publié par Argmax (<https://github.com/argmaxinc/OpenBench/blob/update-diarization-benchmarks/BENCHMARKS.md#diarization-error-rate-der>) , il a obtenu le taux d'erreur le plus faible des six systèmes testés, dont trois services de diarisation cloud, tout en s'exécutant sur un Mac.

NVIDIA fait également état de progrès par rapport à son précédent modèle de diarisation sur chacun des benchmarks qu'elle a testés, et l'écart se creuse lorsque plus de quatre personnes parlent, une situation dans laquelle les étiquettes d'intervenants perdaient souvent en fiabilité.



Graphiques de NVIDIA comparant le taux d'erreur de diarisation selon le nombre d'intervenants sur quatre tests de référence. Le taux d'erreur de Nemotron 3 Diarization reste nettement inférieur à celui du précédent modèle de NVIDIA limité à quatre intervenants, et l'écart est le plus grand avec cinq intervenants ou plus.

Graphique : NVIDIA, extrait de l'annonce de Nemotron 3 Diarization (<https://huggingface.co/blog/nvidia/nemotron-diarization>) . Plus le taux est bas, mieux c'est ; les zones ombrées correspondent à plus de quatre intervenants. NVIDIA publie le modèle sur Hugging Face (<https://huggingface.co/nvidia/Nemotron-3-Diarization>) avec l'intégralité des résultats de son évaluation.

Que change Nemotron 3 Diarization dans Hedy ?

Hedy étiquette déjà les intervenants pendant qu'ils parlent grâce au moteur de reconnaissance vocale Nemotron (/post/nemotron-on-device-speech-engine/) . Ces étiquettes en direct resteront exactement les mêmes.

La nouveauté intervient après la fin d'une session. Les étiquettes en direct doivent être attribuées sur le moment. Hedy réexaminera donc la conversation une fois celle-ci terminée. Nemotron 3 Diarization analysera la session et mettra à jour l'intervenant auquel appartient chaque partie de la transcription.

Ce traitement ne modifiera que l'attribution des passages aux intervenants. Chaque mot de votre transcription et chaque horodatage resteront inchangés. Une fois le traitement terminé, les puces d'intervenants au-dessus de votre transcription (/help/understanding-hedy-transcripts/) fonctionneront comme aujourd'hui : appuyez sur une puce pour donner un nom à l'intervenant, ou fusionnez deux puces qui correspondent en réalité à la même personne.

Nemotron 3 Diarization fonctionne-t-il sur l'appareil ?

Hedy cherche à vous accompagner en réunion en écoutant en temps réel tout en préservant la confidentialité de votre conversation. C'est pourquoi la reconnaissance vocale s'exécute par défaut sur votre appareil, tout comme l'étiquetage des intervenants.

Nemotron 3 Diarization suit le même principe, et nous sommes ravis qu'un modèle aussi performant soit assez compact pour fonctionner sur un téléphone. Il s'exécute sur votre iPhone ou Mac grâce à Core ML d'Apple, et aucun audio n'est envoyé à un service cloud pour déterminer qui parlait. Pour effectuer ce traitement après une session, Hedy conserve l'audio de la session dans un cache temporaire sur l'appareil, puis le supprime. Le guide de la reconnaissance vocale (</help/speech-recognition-providers-in-hedy/>) explique ce cache et le réglage qui le contrôle.

L'étape d'IA qui rédige votre résumé et vos notes relève d'un choix distinct. Vous pouvez également l'exécuter sur votre appareil avec le Traitement IA local (</help/local-ai-processing/>) .

Quels appareils prennent en charge Nemotron 3 Diarization ?

Nemotron 3 Diarization s'exécute sur les iPhone et les Mac avec Apple Silicon compatibles sous Hedy 3.12 ou une version ultérieure. Hedy sur Windows et Android conserve ses étiquettes d'intervenants actuelles.

Hedy 3.12 se déploie progressivement pendant les prochaines 24 heures, alors surveillez sa disponibilité dans Hedy. Une fois la mise à jour installée, ouvrez Paramètres !' Discours et IA et installez la mise à jour du modèle de reconnaissance vocale proposée par Hedy.

Cela s'applique aux sessions capturées avec le moteur de reconnaissance vocale Nemotron, celui qui distingue les intervenants dans Hedy. Si vous utilisez aujourd'hui Whisper ou un fournisseur de reconnaissance vocale cloud, passer à Nemotron dans ce même écran de paramètres vous donnera les étiquettes d'intervenants en direct et après la session.

Questions fréquentes

Qu'est-ce que Nemotron 3 Diarization ?

C'est un modèle de diarisation des intervenants créé par NVIDIA. La diarisation détermine qui a parlé et à quel moment dans un enregistrement. Nemotron 3 Diarization est conçu pour les conversations comme les réunions et les appels, gère les personnes qui parlent en même temps et peut suivre jusqu'à huit intervenants.

Quelle est la précision de Nemotron 3 Diarization ?

Selon NVIDIA, le modèle occupe la première place du classement de diarisation de VoiceArena avec un taux d'erreur de 14,72 %, devant le système suivant à 19,3 %. Dans le benchmark publié par Argmax, il a obtenu le taux d'erreur le plus faible des six systèmes testés, dont trois services de diarisation cloud. Le taux d'erreur de diarisation mesure la proportion de temps de parole attribué au mauvais intervenant, non détecté ou marqué à tort comme de la parole ; plus ce taux est bas, mieux c'est.

Les étiquettes des intervenants changeront-elles pendant une session en direct ?

Non. Les étiquettes affichées pendant une session en direct continueront de fonctionner comme aujourd'hui. Nemotron 3 Diarization intervient après la fin de la session.

Fonctionne-t-il sur mon appareil ?

Oui. Le modèle s'exécute sur votre iPhone ou Mac grâce à Core ML d'Apple. Votre audio n'est pas envoyé à un service cloud pour déterminer qui parlait.

À quelles sessions cela s'appliquera-t-il ?

Aux sessions capturées avec le moteur de reconnaissance vocale Nemotron. Le modèle lit l'audio de la session, qui doit donc être présent sur votre appareil : soit l'enregistrement audio est activé, soit le cache audio temporaire de Nemotron est activé. Ce cache est activé par défaut et supprimé après le traitement.

Connaît-il le nom des personnes ?

Non. Le modèle distingue les voix et les étiquette Speaker 1, Speaker 2, et ainsi de suite. Les noms viennent toujours de Hedy, qui les déduit des présentations, de votre invitation de calendrier et du contexte du sujet, ainsi que des noms que vous saisissez vous-même.

Quels appareils le prennent en charge ?

Les iPhone et les Mac avec Apple Silicon compatibles exécutant Hedy 3.12 ou une version ultérieure. Hedy sur Windows et Android conserve ses étiquettes d'intervenants actuelles.

Hedy AI · Coaching IA en direct pour les conversations importantes

Essayez Hedy gratuitement: <https://www.hedy.ai/fr/downloads/>

<https://www.hedy.ai/fr/post/nemotron-3-diarization-on-device/>