

À PROPOS DE L'AUTRICE



Juthika Thakur, M.D.

La D^{re} Juthika Thakur a obtenu son baccalauréat en sciences médicales à l'Université Western et un diplôme de commerce, avec distinction, à la *Richard Ivey School of Business* en 2011. Elle a ensuite obtenu son diplôme de docteure en médecine à la *Michael G. DeGroot School of Medicine* de l'Université McMaster, puis a effectué son internat en dermatologie à l'Université de Toronto, où elle a été corésidente en cheffe pendant sa dernière année. Depuis, la D^{re} Thakur a rédigé et présenté ses travaux de recherche lors de plusieurs conférences nationales et internationales, notamment celles de l'Association canadienne de dermatologie, de l'*European Academy of Dermatology and Venereology* et du Congrès mondial de dermatologie. Elle a également publié des articles dans la revue *Ivey Business Review*. Elle s'intéresse à l'interface entre la santé en ligne, l'apprentissage automatique et la dermatologie.

Guide d'initiation pour les dermatologues : Utiliser l'intelligence artificielle ambiante dans votre pratique

Juthika Thakur, M.D.

Introduction

La reconnaissance automatique de la parole (RAP) est la technologie de base qui permet de convertir la langue parlée en texte écrit. Parmi ses points forts figurent la capacité à interpréter les voix, à identifier les différents intervenants dans une conversation et à suivre le déroulement d'un dialogue. Avec l'aide de formateurs humains, les systèmes de RAP peuvent encore améliorer et optimiser leurs performances à partir des rétroactions. Cependant, leur efficacité peut être limitée par des facteurs tels que des environnements bruyants, un mauvais positionnement des microphones ou des variations de dialectes et d'accents. Voir l'**Annexe 1** pour les définitions des principaux termes liés à l'IA.

Les grands modèles de langage (GML) visent à combler les lacunes laissées par les systèmes de RAP en utilisant le traitement automatique du langage naturel (TALN) pour compléter les éléments manquants d'une conversation. En tant que forme d'IA générative, un GML prédit le mot suivant dans une séquence à partir des schémas appris dans ses données d'entraînement (**Figure 1**). L'objectif d'un GML est d'imiter le langage humain en identifiant quels mots sont susceptibles de suivre les autres selon le contexte fourni dans l'invite et l'entraînement antérieur. Les GML peuvent générer du contenu original et optimiser leurs sorties à partir d'une entrée humaine. Cependant, comme ils fonctionnent par prédiction de séquences de mots, ils ne savent pas si leurs

Principaux termes et définitions liés à l'IA
Intelligence artificielle : Domaine de l'informatique visant à créer des systèmes capables d'exécuter des tâches nécessitant normalement l'intelligence humaine, comme le raisonnement, la résolution de problèmes, l'apprentissage et la compréhension du langage.
Apprentissage automatique : Sous-domaine de l'IA dans lequel les ordinateurs apprennent des schémas à partir de données pour prendre des décisions ou faire des prédictions, sans être explicitement programmés pour chaque tâche.
Intelligence augmentée : Approche collaborative dans le cadre de laquelle humains et systèmes d'IA travaillent ensemble pour améliorer la prise de décision et la performance humaines, plutôt que de remplacer l'intelligence humaine.
IA agentique : Systèmes d'IA conçus pour fonctionner de manière autonome et poursuivre des objectifs ou effectuer des actions de façon proactive, avec un certain degré de prise de décision ou d'initiative.
Grands modèles de langage (GML) : Modèles d'IA entraînés sur d'immenses corpus de textes afin de comprendre et générer un langage ressemblant au langage humain. Exemples : ChatGPT, Claude et Gemini.
Traitement automatique du langage naturel (TALN) : Branche de l'IA visant à permettre aux machines de comprendre, d'interpréter et de générer le langage humain, écrit comme oral.
Hallucinations : Situation dans laquelle un modèle d'IA génère une information qui paraît plausible, mais qui est en réalité fausse ou non fondée sur des données réelles.
Apprentissage supervisé : Type d'apprentissage automatique où le modèle est entraîné sur des données étiquetées (paires entrée-sortie), ce qui lui permet d'apprendre à faire des prédictions ou des classifications.
Apprentissage non supervisé : Technique d'apprentissage automatique dans laquelle le modèle est entraîné sur des données sans résultats étiquetés, souvent utilisée pour trouver des schémas ou créer des regroupements (comme l'agrégation ou « clustering »).
Biais : Erreurs systématiques dans les sorties de l'IA causées par des données d'entraînement déséquilibrées ou défectueuses, par des choix de conception ou par des inégalités sociales, d'où des résultats injustes ou inexacts.
Explicabilité et interprétabilité : Degré selon lequel les humains peuvent comprendre le fonctionnement d'un modèle d'IA, y compris la manière dont il prend ses décisions ou fait ses prédictions. Ces notions sont fondamentales pour instaurer la confiance et assurer la transparence des systèmes d'IA.

Annexe 1. Principaux termes et définitions liés à l'IA; avec l'aimable autorisation du Juthika Thakur, M.D.

sorties sont vraies ou fausses. C'est pourquoi, en réponse aux invites, les GML peuvent produire des réponses inexacts ou inventées, communément appelées « *hallucinations* ».

Les scribes d'IA ambiante, qui utilisent l'apprentissage automatique pour traiter les conversations en temps réel, présentent un potentiel prometteur pour : réduire la charge liée à la documentation, améliorer la qualité des interactions entre médecin et patient et soutenir les cliniciens dans leur pratique quotidienne. Lorsqu'ils sont combinés, les GML et les systèmes de RAP peuvent compenser mutuellement leurs limites, ce qui donne naissance à un scribe d'IA ambiante plus efficace et adapté aux milieux cliniques (**Figure 2**).

Exactitude du dossier médical électronique généré par le médecin ou par une IA ambiante

La documentation clinique ne reflète pas toujours fidèlement le fond de la consultation avec le patient, avec des omissions qui ne reflètent pas nécessairement la nature ou l'intention de l'échange. Bien qu'aucune étude n'ait évalué l'utilisation des scribes IA en dermatologie, les résultats issus d'autres spécialités semblent généralisables. Une étude portant sur 36 médecins dans un même centre a comparé les notes de consultation écrites à des enregistrements audio dissimulés de consultations non annoncées. Les résultats ont montré que 90 % des dossiers des patients contenaient au moins une inexactitude, incluant des omissions ou des ajouts ne reflétant

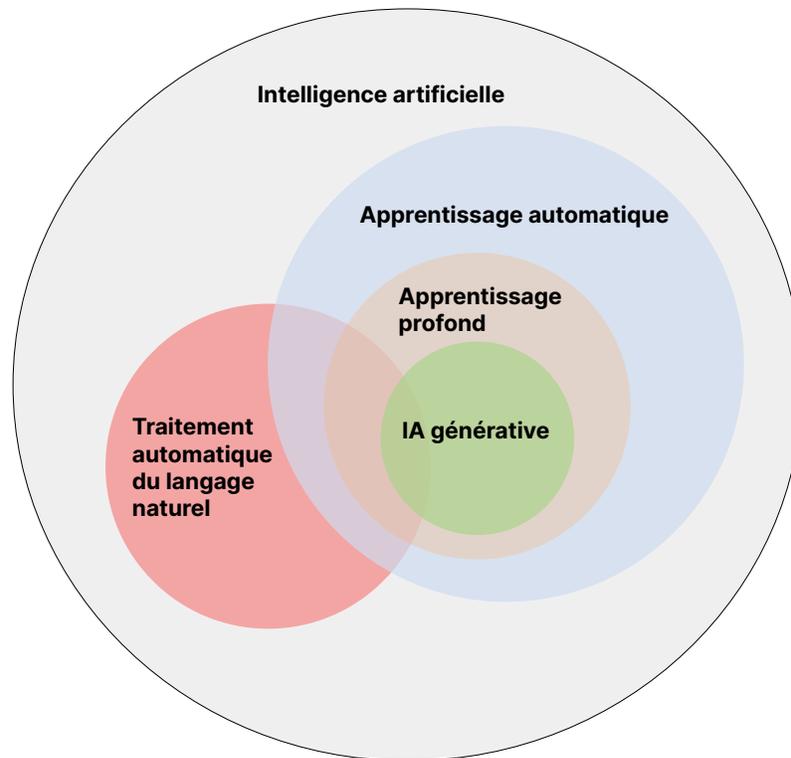


Figure 1. Diagramme de Venn illustrant divers sous-domaines de l'IA; adapté de Rishabh Misra, 2024.⁷

Reconnaissance automatique de la parole	Grands modèles de langage
<p>Capacités de base</p> <ul style="list-style-type: none"> ✓ Capturer, reconnaître et interpréter des voix ✓ Identifier les différents interlocuteurs dans une conversation ✓ Suivre les schémas conversationnels <p>Limites de base</p> <ul style="list-style-type: none"> → Sensibilité aux conditions environnementales (placement des micros, bruit ambiant) → Difficultés potentielles avec les termes, accents ou dialectes absents des données d'entraînement → Difficultés potentielles à traiter les conversations fragmentées et non linéaires 	<p>Capacités de base</p> <ul style="list-style-type: none"> ✓ Identifier les mots susceptibles d'apparaître ensemble à partir des données d'entraînement et du contexte fourni par l'invite ✓ Produire des contenus réalistes et convaincants ✓ Générer du contenu original ✓ Apprendre à partir des corrections fournies par les utilisateurs ou les entraîneurs humains ✓ Améliorer et optimiser les performances au fil du temps <p>Limites de base</p> <ul style="list-style-type: none"> → La qualité des données d'entrée détermine celle des données de sortie → Non conçu pour distinguer le vrai du faux → Les grands modèles de langage improvisent et hallucinent → Futures difficultés avec des concepts absents des données d'entraînement → Peut reproduire les biais présents dans les données d'entraînement

Figure 2. Forces et faiblesses de la reconnaissance automatique de la parole (RAP) et des grands modèles de langage (GMD); adapté de Information Services Centre of Effective Practice. (Anne Dabrowski, n.d.)⁸

Caractéristique	Description de la note idéale
Exacte	La note est véridique. Elle ne contient aucune information incorrecte.
Exhaustive	La note est complète, sans omission, et documente tous les éléments importants pour le patient.
Utile	La note est extrêmement pertinente, apportant une information et/ou une analyse de grande valeur.
Bien structurée	La note est bien construite et organisée de manière à faciliter la compréhension du parcours clinique du patient.
Compréhensible	La note est compréhensible, sans ambiguïté ni passages difficiles à comprendre.
Concise	La note est brève, va à l'essentiel et ne contient pas de redondance.
Synthétisée	La note reflète la compréhension de l'état du patient par le scribe IA et sa capacité à élaborer un plan de soins.
Présente une cohérence interne	Aucune partie de la note ne contredit ou n'ignore une autre partie.
Exempte d'hallucinations	La note ne contient aucune hallucination, uniquement des informations vérifiables à partir de la transcription.
Exempte de biais	La note est objective, contenant des informations vérifiables à partir de la transcription, sans influence des caractéristiques du patient ou de la consultation.

Annexe 2. Éléments d'un résumé clinique idéal; *adapté de Tierney et coll., 2024.*³

Version modifiée du Physician Documentation Quality Instrument (Instrument de qualité de la documentation du médecin), adaptée à l'évaluation des sorties produites par un scribe IA.

La valeur maximale est de 50, et chaque domaine est évalué sur une échelle de 5 points, où 1 signifie « pas du tout » et 5 « tout à fait ».

pas la nature de l'échange capturé dans les enregistrements audio dissimulés.¹ Dans une autre étude menée auprès de 136 815 patients ayant consulté leurs notes de consultation externe, 20 % ont détecté une erreur, et 40 % d'entre eux considéraient l'erreur comme grave.² Les erreurs les plus fréquentes : diagnostics, antécédents médicaux, médicaments, examens cliniques, et notes attribuées au mauvais patient.² De nombreux utilisateurs de l'IA se plaignent de l'exactitude diagnostique des algorithmes d'apprentissage automatique et des GML. Malheureusement, le dossier du patient ne reflète souvent pas fidèlement la rencontre entre le patient et le médecin. Dans un programme pilote régional ayant déployé des solutions de scribes médicaux IA auprès de 10 000 médecins, les expériences des patients et des médecins se sont améliorées.³

Une analyse aléatoire de 35 notes cliniques dans plusieurs spécialités a révélé que plus de 90 % respectaient des critères tels que l'absence d'hallucinations, la concision et l'exactitude (**Annexe 2**). Bien que cette étude ait montré de légères améliorations du temps consacré à la rédaction des notes médicales,³ d'autres études indiquent une réduction de 20,4 % du temps consacré à la rédaction.⁴ L'entraînement spécialisé par domaine, avec intervention humaine (« *humans in the loop* »), peut limiter les erreurs de transcription. Par exemple, grâce aux rétroactions et en apportant des ajustements au module d'IA ambiante, les sorties peuvent être ajustées pour refléter le style propre à chaque clinicien. Les modèles non entraînés sur des données de santé ont tendance à mal interpréter la terminologie médicale, à confondre les noms de médicaments,

à mal saisir la complexité des échanges médicaux et, de manière générale, à rencontrer des difficultés en raison d'une technologie inadéquate et d'environnements bruyants.

Boîte noire, explicabilité et interprétabilité

Il reste difficile d'évaluer l'exactitude diagnostique des GML, en raison de la nature opaque de leurs ensembles de données d'entraînement, qui sont souvent exclusifs et non accessibles aux chercheurs. Ce manque de transparence limite la capacité à identifier, prédire ou réduire les biais et angles morts potentiels dans les sorties des modèles. L'explicabilité désigne la mesure dans laquelle les processus à l'origine des sorties d'une IA peuvent être clairement compris ou justifiés. Des mesures ont été mises en place pour fournir un certain degré d'explicabilité. Par exemple, certains systèmes intègrent des fonctionnalités permettant d'ajouter, dans les notes générées, des « citations » renvoyant à des segments précis de la transcription audio que le scribe IA a résumés sous forme de phrases. Cela peut contribuer à réduire les hallucinations, mais exige encore un important travail de relecture et de correction de la part des médecins. Des recherches récentes sur l'usage de l'IA par les cliniciens mettent en évidence un risque réel de biais d'automatisation, pouvant entraîner des relectures moins rigoureuses et des erreurs non détectées.⁵ En définitive, les fournisseurs de solutions d'IA ambiante doivent s'efforcer de développer des technologies explicables pour l'ensemble des parties prenantes, notamment les patients et les médecins, lorsqu'elles sont utilisées comme scribes IA dans le cadre de consultations médicales.

Confidentialité et consentement éclairé dans le cadre de l'usage de l'IA ambiante

Les cliniciens ont des obligations éthiques et légales d'obtenir le consentement éclairé des patients. Les patients peuvent souhaiter obtenir des précisions sur la durée de conservation des enregistrements, les personnes autorisées à y accéder ou à consulter le résumé généré par l'IA, si et comment le médecin procédera à la relecture, si le matériel recueilli est utilisé pour entraîner une IA générative, ainsi que sur le risque potentiel de réidentification. Un défi majeur réside dans le fait que, une fois le consentement donné, les patients ne peuvent généralement pas le révoquer pour empêcher l'usage futur de leurs données dans le cadre de l'entraînement des modèles. Ainsi, le consentement éclairé devrait inclure une discussion sur la question de savoir si et comment les données seront anonymisées et utilisées afin d'améliorer l'algorithme. Pour limiter ces risques, les cliniciens peuvent choisir un scribe IA ambiante qui héberge les données sur des serveurs locaux et évite l'utilisation indéfinie des données patients.

Défis liés à l'archivage des dossiers

Bien que l'entrée finale dans le dossier doive être conservée selon les règles de conservation, les lois et règlements ne précisent généralement pas si les enregistrements audio doivent également être inclus dans le dossier patient. Lorsque les enregistrements sont destinés à être détruits plutôt que conservés, il est essentiel d'adopter une politique définissant le moment et le processus de destruction, en veillant à ce que le dossier patient ait été mis à jour correctement au préalable.⁶ Certaines instances réglementaires exigent un formulaire de consentement supplémentaire pour l'enregistrement des consultations et, dans certaines provinces, les lois sur la protection de la vie privée imposent un consentement écrit pour tout enregistrement.⁶ Les médecins doivent donc examiner attentivement les lois sur la protection de la vie privée et les réglementations provinciales applicables avant d'intégrer des technologies de scribes IA dans leurs pratiques.

Résumé

Les méthodes d'évaluation approfondie de la qualité et de la sécurité des technologies d'IA, y compris les GML, ne sont pas encore pleinement établies. Tant les algorithmes que les cadres réglementaires continuant d'évoluer, il sera nécessaire de mettre en place des référentiels de comparaison, des évaluations et une surveillance continus. Enfin, l'adoption et les usages de ces technologies évolueront probablement avec l'émergence de nouveaux groupes d'utilisateurs et de nouveaux domaines d'application.

Autrice correspondante

Juthika Thakur, M.D.

Courriel : juthika.thakur@medportal.ca

Divulgations des liens financiers

J.T. : Aucun déclaré.

Références

- Weiner SJ, Wang S, Kelly B, Sharma G, Schwartz A. How accurate is the medical record? A comparison of the physician's note with a concealed audio recording in unannounced standardized patient encounters. *J Am Med Inform Assoc.* 2020;27(5):770–775. doi:10.1093/jamia/ocaa027
- Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. *JAMA Netw Open.* 2020;3(6):e205867. doi:10.1001/jamanetworkopen.2020.5867
- Tierney AA, Gayre G, Hoberman B, Mattern B, Balleca M, Kipnis P, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. *NEJM Catal Innov Care Deliv.* 2024;5(3). doi:10.1056/CAT.23.0404
- Duggan MJ, Gervase J, Schoenbaum A, Hanson W, Howell JT, 3rd, Sheinberg M, et al. Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. *JAMA Netw Open.* 2025;8(2):e2460637. doi:10.1001/jamanetworkopen.2024.60637
- Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. *Jama.* 2023;330(23):2255–2257. doi:10.1001/jama.2023.22557
- CMPA. AI Scribes: answers to frequently asked questions Ottawa, Ontario, Canada: CMPA; 2023: [Cited June 6 2025] Available from: <https://www.cmpa-acpm.ca/en/advice-publications/browse-articles/2023/ai-scribes-answers-to-frequently-asked-questions>.
- Misra R. What is generative AI? : Medium; 2024: [Cited June 6 2025] Available from: <https://medium.com/analytics-vidhya/what-is-generative-ai-9c34cfa0fd6b>.
- Dabrowski A. OMD Educates—AI 101: Intro to AI and AI Scribe [Video Recording]: OntarioMD; 2024: [Cited March 22 2025] Available from: <https://vimeo.com/925848578>.