

Les impacts de l'utilisation des outils d'intelligence artificielle pour la documentation clinique sur les professionnels, leur travail et les systèmes de santé

Revue rapide

Une production de l'Unité d'évaluation des technologies et des modes d'intervention (UETMI), Direction de l'enseignement universitaire et de la recherche (DEUR), CIUSSS du Centre-Sud-de-l'Île-de-Montréal (CCSMTL)

Roua Walha, conseillère scientifique, UETMI, DEUR, CCSMTL

Akram Djouini, conseiller scientifique, UETMI, DEUR, CCSMTL

Fannie Tremblay-Racine, bibliothécaire et responsable des

Bibliothèques en déficience physique, DEUR, CCSMTL

Novembre 2025



Auteurs

Roua Walha, Ph. D., conseillère scientifique, UETMI, DEUR, CCSMTL

Akram Djouini, M.D., M. Sc., conseiller scientifique, UETMI, DEUR, CCSMTL

Fannie Tremblay-Racine, B.A., M.S.I., bibliothécaire et responsable des Bibliothèques en déficience physique, DEUR, CCSMTL

Soutien au repérage des documents

Noémie Delarsobil, technicienne en documentation, Bibliothèques en déficience physique, DEUR, CCSMTL

Coordination scientifique

Isabelle Linteau, Ph.D., coordonnatrice scientifique, UETMI, DEUR, CCSMTL

Gestionnaires du mandat

Catherine Safianyk, chef de service des Bibliothèques et de l'UETMI, DEUR, CCSMTL

Nathalie Dupont, directrice adjointe – Enseignement, pour les services Stages, ETMI, Bibliothèques et Expérience stagiaire, DEUR, CCSMTL

Demandeur

Geneviève Gagnon, adjointe au directeur, Programmes transversaux et guichet d'accès, Direction des programmes en déficiences et de la réadaptation physique, CISSS des Laurentides

Représentante du demandeur

Marilyn Larente, chef de programme, Réadaptation hospitalière et déficience physique Nord, Direction des programmes en déficiences et de la réadaptation physique, CISSS des Laurentides

Collaboratrice au contenu clinique

Marie-Michèle Lapointe, spécialiste en activités cliniques, Direction des programmes en déficiences et de la réadaptation physique, CISSS des Laurentides

Comité de priorisation

Manon Parisien, chef de service, IURDPM, DEUR, CCSMTL

Chantal Robillard, chef d'équipe, Direction des Affaires académiques et de l'éthique de la recherche, Centre de réadaptation Lethbridge-Layton-Mackay, CIUSSS du Centre-Ouest-de-l'Île-de-Montréal

Stéphanie Laurin, adjointe à la directrice par intérim, Direction des programmes DI-TSA-DP, CISSS de Laval

Cloé Rodrigue, directrice de la recherche, CISSS de la Montérégie-Centre

Renée Rancourt, chef en réadaptation en déficience auditive et en déficience visuelle, Direction des programmes DI-TSA-DP, CISSS de Lanaudière (jusqu'en août 2024)

Lynda Cloutier, APPR, Direction des services multidisciplinaires de l'enseignement et de la recherche, CISSS des Laurentides (jusqu'en décembre 2024)

Isabelle Linteau, coordonnatrice scientifique, UETMI, DEUR, CCSMTL

Catherine Safianyk, chef de service des Bibliothèques et de l'UETMI, DEUR, CCSMTL

Conception graphique

Équipe Diffusion des connaissances, DEUR, CCSMTL

Révision linguistique

Nada Ndaya, agente administrative, UETMI, DEUR

Responsabilité

L'UETMI sous la direction de la DEUR du CCSMTL assume la responsabilité de la forme et du contenu définitifs du présent document. Les conclusions et constats formulés ne reflètent pas nécessairement les opinions du CCSMTL ou autres personnes consultées aux fins du présent dossier.

Pour citer ce document

Walha, R., Djouini, A., et Tremblay-Racine, F. (2025). Les impacts de l'utilisation des outils d'intelligence artificielle pour la documentation clinique sur les professionnels, leur travail et les systèmes de santé. Revue rapide. Montréal, Qc: Unité d'évaluation des technologies et des modes d'intervention, Direction de l'enseignement universitaire et de la recherche, CIUSSS du Centre-Sud-de-l'Île-de-Montréal, 67 p.

Note

Dans ce document, l'emploi du masculin générique désigne aussi bien les femmes que les hommes et est utilisé dans le seul but d'alléger le texte.

Ce document est disponible à ccsmtl-mission-universitaire.ca/fr/uetmi

Correspondance

Toutes correspondances doivent être adressées à <u>uetmi.ccsmtl@ssss.gouv.qc.ca</u>

©CIUSSS du Centre-Sud-de-l'Île-de-Montréal, 2025

ISBN (En ligne) 978-2-555-02570-7

Dépôt légal Bibliothèque et Archives nationales du Québec, 2025 Bibliothèque et Archives Canada, 2025 Cette revue rapide a été réalisée dans le cadre d'un partenariat avec les établissements membres du <u>Centre de recherche interdisciplinaire en réadaptation du Montréal métropolitain</u> (<u>CRIR</u>). Plus précisément, le CRIR regroupe quatre établissements membres gestionnaires :

- le CIUSSS du Centre-Sud-de-l'Île-de-Montréal, qui comprend l'Institut universitaire sur la réadaptation en déficience physique de Montréal (IURDPM);
- le CIUSSS du Centre-Ouest-de-l'Île-de-Montréal, qui comprend le Centre de réadaptation Lethbridge-Layton-Mackay;
- le CISSS de Laval, qui comprend l'Hôpital juif de réadaptation ;
- le CISSS de la Montérégie-Centre, qui comprend l'Institut Nazareth et Louis-Braille.

Le CRIR compte également deux établissements membres gestionnaires associés : le CISSS de Lanaudière et le CISSS des Laurentides. Ensemble, ces six établissements offrent des services pour tous les types de déficience physique (motrice, auditive, visuelle, du langage et de la communication) auprès de personnes de tous âges.

PRINCIPAUX CONSTATS

Quels sont les impacts de l'utilisation des outils d'intelligence artificielle pour la documentation clinique sur les professionnels, leur travail et les systèmes de santé?

Cette revue rapide synthétise les connaissances actuelles concernant les effets de l'utilisation des outils d'intelligence artificielle (IA) pour la documentation clinique sur les professionnels, leur travail et les systèmes de santé. Vingt (20) études publiées au cours des cinq dernières années ont été incluses. La majorité a été réalisée aux États-Unis et portait sur des cliniciens de première ligne ainsi que sur des médecins spécialistes, tandis qu'une seule étude s'est intéressée spécifiquement aux professionnels de la réadaptation. Les outils recensés reposaient sur divers modèles d'IA (reconnaissance vocale, traitement automatique du langage naturel, modèles de langage de grande taille) et étaient principalement utilisés pour générer automatiquement des notes cliniques et des résumés de consultation.

1. Impacts sur les professionnels de la santé et leur travail

I.I. Impacts sur les professionnels de la santé

Résultats prometteurs

- Réduction de la charge de travail liée à la documentation:
 Ce bénéfice semble plus marqué chez les cliniciens présentant une meilleure littératie numérique et après une période de familiarisation avec l'outil;
- Amélioration de la satisfaction professionnelle et de l'équilibre travail-vie personnelle;
- Renforcement de l'attention et de l'engagement envers les patients, avec une réduction des distractions durant les consultations.

Résultats variables

• Les effets sur l'épuisement professionnel sont variables selon les outils de mesure utilisés et les dimensions évaluées.

I.II. Impacts sur le travail des professionnels

Résultats prometteurs

- Réduction du temps de documentation ;
- Diminution du temps de travail en dehors des heures régulières ;

Résultats non concluants en raison du manque de données

- La durée des consultations ;
- Le flux de travail.

II. Impacts sur les systèmes de santé

Très peu d'études se sont intéressées aux impacts des outils d'IA sur les systèmes de santé. Celles recensées portent principalement sur les dimensions de l'efficience et de la productivité. En revanche, d'autres dimensions importantes en rapport avec la performance des systèmes, telles que les temps d'attente, la qualité des soins ou les coûts associés, demeurent largement inexplorés.

Résultats prometteurs

Amélioration de la productivité globale des cliniciens.

Résultats non concluants en raison du manque de données

L'efficience.

Précautions quant à l'interprétation

- La robustesse des devis utilisés est généralement faible à modérée : la plupart sont de type pré-post sans groupe témoin ou avec un devis intra-sujet.
- La qualité méthodologique de la grande majorité des études est faible : échantillons restreints et non représentatifs, avec un contrôle insuffisant des facteurs de confusion.
- Les résultats reposent majoritairement sur des mesures subjectives (ex : questionnaires auto-rapportés) et des instruments de mesure non validés dans plusieurs études.
- Plusieurs études ont été menées dans des contextes expérimentaux contrôlés, ce qui limite la généralisation des résultats aux milieux cliniques réels.

TABLE DES MATIÈRES

ABRÉVIATIO	NS	9
GLOSSAIRE		10
INTRODUCTI	ON	12
MANDAT		13
MÉTHODOLO	OGIE	13
Qu'est-ce	qu'une revue rapide ?	13
Question o	lécisionnelle et questions d'évaluation	13
PICOTS et	critères d'inclusion et d'exclusion	14
Recherche	documentaire	15
Sélection o	des documents	15
Évaluation	de la qualité méthodologique des documents retenus	16
Extraction	et synthèse des données	16
RÉSULTATS		17
I. Impa	cts sur les professionnels de la santé et leur travail	18
1.1.	mpacts sur les professionnels de la santé	18
a.	Charge de travail liée à la documentation	18
b.	Épuisement professionnel	19
C.	Satisfaction professionnelle, bien-être et équilibre travail-vie personnelle	19
d.	Engagement des professionnels	20
1.11.	mpacts sur le travail des professionnels	20
a.	Temps de documentation	20
b.	Durée des consultations	21
c.	Temps de travail en dehors des heures régulières	21
d.	Flux de travail	22
Quali	tés méthodologiques des études	22
SOMI	MAIRE	23
II. Impa	cts sur les systèmes de santé	24
a.	Efficience	24
b.	Productivité	25

Qualité méthodologique des résultats	25
SOMMAIRE	26
CONCLUSION	27
Forces et limites de la revue rapide	27
RÉFÉRENCES	28
ANNEXE 1. RECHERCHE DOCUMENTAIRE	32
Bases de données scientifiques consultées	32
Stratégies de recherche dans les bases de données	33
ANNEXE 2. DIAGRAMME DE SÉLECTION	37
ANNEXE 3. ÉVALUATION DE LA QUALITÉ DES ÉTUDES (MMAT)	38
ANNEXE 4. CARACTÉRISTIQUES DES ÉTUDES	41
ANNEXE 5. VARIABLES D'INTÉRÊT ET OUTILS DE MESURE	52
ANNEXE 6. ANALYSES STATISTIQUES, LIMITES MÉTHODOLOGIQUES ET RÉSULTATS	56

Tableau 1 – Critères de sélection des documents basés sur le modèle PICOTS et autres critères.....14

ABRÉVIATIONS

CCSMTL CIUSSS du Centre-Sud-de-l'Île-de-Montréal

CISSS Centres intégrés de santé et de services sociaux

CIUSSS Centres intégrés universitaires de santé et de services sociaux

DEUR Direction de l'enseignement universitaire et de la recherche

DME Dossier médical électronique

ECR Essai contrôlé randomisé

GI Groupe d'intervention

GT Groupe témoin

IA Intelligence artificielle

MMAT Mixed Methods Appraisal Tool

NASA-TLX NASA Task Load Index

N/A Non applicable

N.B. Notez bien

OLBI Oldenburg Burnout Inventory

S.O. Sans objet

PDQI-9 Documentation Quality Instrument

PFI Professional Fulfillment Index

PTL Physician Task Load

UETMI Unité d'évaluation des technologies et des modes d'intervention

VBC Value-Based Care

wRVU Work Relative Value Units

GLOSSAIRE¹

Documentation clinique	La documentation clinique désigne toute note manuelle ou électronique inscrite au dossier d'un patient par un clinicien, en lien avec son état de santé ou son traitement. Source : Canadian Institute for Health Information
Dossier médical électronique	Un dossier médical électronique (DME) est un dossier médical informatique tenu par les cliniciens pour chacun de leurs patients. Il contient des renseignements détaillés, tels que les données démographiques, les antécédents médicaux et pharmacologiques, ainsi que l'information diagnostique comme les résultats de laboratoire et d'imagerie.
	Source : Le Collège des médecins de famille du Canada
Facteur de confusion	Variable externe distribuée différemment dans la population étudiée et qui fausse la relation entre la variable indépendante, à laquelle elle est associée sans en être la conséquence, et la variable dépendante, avec laquelle elle entretient une relation causale. Source: Office québecois de la langue française
IA ambiante	L'IA ambiante désigne des systèmes intelligents intégrés à l'environnement, capables de fonctionner de manière discrète tout en recueillant des données issues de multiples sources. Leur objectif est d'offrir une assistance fluide, proactive et adaptée au contexte, avec une intervention humaine minimale. Par exemple, l'IA ambiante peut être utilisée pour transcrire automatiquement les échanges entre le clinicien et son patient. Source: OpenTrain Al
IA générative	Branche de l'IA mettant en œuvre des modèles génératifs, qui vise à produire des contenus textuels, graphiques ou audiovisuels. Source : Office québecois de la langue française
Modèle de langage de grande taille	Un modèle d'IA de très grande taille entraîné sur d'énormes corpus de données généralistes et conçu dans le but d'effectuer une large gamme de tâches linguistiques.

¹ Les définitions suivantes sont tirées partiellement ou intégralement des sources citées en référence.

	Source : Office québecois de la langue française
Outils d'IA pour la documentation clinique	Dans le cadre de la présente revue, tout logiciel ou application utilisant l'IA pour assister les professionnels de santé dans la création, l'organisation et la synthèse de notes cliniques, comptes rendus ou résumés.

INTRODUCTION

La documentation clinique constitue un pilier fondamental de la pratique médicale, assurant la continuité, la qualité et la traçabilité des soins. Toutefois, elle représente également une source majeure de charge administrative pour les professionnels de la santé. Une étude menée aux États-Unis rapporte que ces derniers consacrent en moyenne deux heures supplémentaires par jour, en dehors de leurs heures régulières de travail, à des tâches de documentation (Gaffney et al., 2022). Cette surcharge administrative est l'une des principales causes d'épuisement professionnel chez les cliniciens, tout en réduisant le temps consacré aux soins directs aux patients (Muhiyaddin et al., 2022).

L'utilisation d'outils d'intelligence artificielle (IA) dans le domaine de la santé a connu une forte expansion au cours des dernières années, en particulier dans la documentation clinique, où elle représente une innovation prometteuse (Bongurala et al., 2024). Ces technologies reposent sur des modèles de reconnaissance vocale, de language de grande taille et d'apprentissage machine pour assister les cliniciens dans la rédaction de différents documents tels que les notes, les résumés ou les lettres médicales (Bracken et al., 2025). Elles peuvent fonctionner selon différents modes, tels que la dictée assistée, où le clinicien dicte son texte et l'IA le transcrit automatiquement, la génération de résumés concis et structurés à partir de notes déjà produites, ou encore la rédaction ambiante, où l'outil capte et analyse la conversation entre le clinicien et le patient afin de générer un compte rendu prérempli (Bongurala et al., 2024).

L'implantation de ces outils vise à optimiser le processus de documentation. Dans ce contexte d'évolution rapide, il devient donc essentiel de mieux comprendre leurs retombées sur les pratiques et l'organisation des soins.

C'est dans cette perspective que la présente revue rapide de la littérature vise à synthétiser les connaissances scientifiques récentes portant sur l'utilisation des outils d'IA pour la documentation clinique. Plus précisément, elle cherche à examiner leurs effets sur les professionnels de la santé, leur travail et les systèmes de santé.

MANDAT

Dans un contexte de vieillissement démographique et de hausse de la demande en services de réadaptation, les systèmes de santé sont soumis à des pressions croissantes. Les professionnels doivent composer avec des charges administratives lourdes tout en répondant à des exigences de performance toujours plus élevées. La rédaction de documents cliniques représente une part importante de leurs tâches, soulevant la question de l'optimisation de ce processus. Dans cette optique, la Direction des programmes en déficiences et de la réadaptation physique du CISSS des Laurentides souhaite évaluer la pertinence du recours à des technologies de rédaction assistée par l'intelligence artificielle afin de soutenir les intervenants dans leurs activités cliniques. À cette fin, elle a mandaté l'UETMI de la DEUR du CIUSSS du Centre-Sud-de-l'Île-de-Montréal pour réaliser une revue rapide de la littérature sur ce sujet.

MÉTHODOLOGIE

Qu'est-ce qu'une revue rapide?

La revue rapide est une approche scientifique transparente visant à synthétiser les connaissances issues de la littérature, constituant un compromis entre l'exhaustivité d'une revue systématique et une analyse plus ciblée, grâce à certaines adaptations méthodologiques (Hamel et al., 2021; Tricco et al., 2020). Les lignes directrices établies dans ce domaine ont été suivies, notamment en limitant les biais potentiels liés à la sélection des études grâce à un tri interjuges et en évaluant rigoureusement la qualité méthodologique des travaux inclus (Garritty et al., 2021; Hamel et al., 2020; Tricco et al., 2016).

Question décisionnelle et questions d'évaluation

Dans le cadre de cette revue rapide, la question décisionnelle est la suivante :

Quels sont les impacts de l'utilisation des outils d'IA pour la documentation clinique sur les professionnels de la santé, leur travail et les systèmes de santé?

Partant de cette grande question, deux questions d'évaluation ont été formulées :

- **Q1.** Quels sont les impacts de l'utilisation des outils d'IA pour la documentation clinique sur les professionnels de la santé et leur travail?
- **Q2.** Quels sont les impacts de l'utilisation des outils d'IA pour la documentation clinique sur les systèmes de santé ?

Pour y répondre, un examen de la littérature scientifique a été réalisé.

PICOTS et critères d'inclusion et d'exclusion

Le modèle PICOTS (Population – Intervention – Comparateur – *Outcomes* – Temporalité – *Setting*) a été utilisé pour formuler les questions d'évaluation ci-dessus, orienter la recherche documentaire et préciser les critères de sélection des documents. D'autres critères de sélection concernant les devis des études, les types de publications, les langues, les pays et les années de publication ont été ajoutés (voir Tableau 1 ci-dessous).

Tableau 1 – Critères de sélection des documents basés sur le modèle PICOTS et autres critères

CRITÈRES	INCLUSION	EXCLUSION
Population	Professionnels de la santé	Autres professionnels (p. ex. personnel administratif)
Intervention	Outils d'IA conçus pour assister la rédaction professionnelle: Transcription des visites Génération de résumé Rédaction de notes de consultation et de suivi	 Outils d'IA utilisés pour Établir un diagnostic Rédiger des rapports à partir d'images ou de vidéos Élaborer un plan de traitement Interpréter/analyser et évaluer des données cliniques Codifier les maladies/syndromes
Comparateur	Avec ou sans comparateur	S.O.
<i>Outcomes</i> (résultats d'intérêt)	 Q1 Impacts sur les professionnels: stress, charge de travail, épuisement professionnel, etc. Impacts sur le travail des professionnels: durée des consultations, temps de documentation, temps en dehors des heures régulières, etc. Q2 Impacts sur les systèmes de santé: Productivité, efficience, temps d'attente, qualité des soins, coûts, etc. 	 Qualité de la documentation : précision, erreurs, etc. Enjeux éthiques : confidentialité et sécurité des données Résultats relatifs à la dimension coût-efficacité
Temporalité	S.O.	S.O.
Setting (milieu)	Milieux internes et externes	S.O.

AUTRES CRITÈRES	RITÈRES DE SÉLECTION		
Devis	Aucune restriction	S.O.	
Types de publication	Articles de revue scientifique	Éditoriaux, commentaires, résumés d'article ou de conférence, présentations PowerPoint, protocoles de recherche, chapitres de livre, mémoires de maîtrise et thèses de doctorat	
Langues	Anglais et français	Autres langues	
Pays	Pays membres de l'OCDE	Pays qui ne sont pas membres de l'OCDE	
Années	Publications des cinq dernières années (2020-2025)	Publications datant de plus de cinq ans (2019 et moins)	

S.O.: sans objet

Recherche documentaire

La recherche documentaire a été menée en collaboration avec une bibliothécaire spécialisée (FTR) en mai 2025. Trois bases de données bibliographiques, *Medline (Ovid), All EBM Reviews (Ovid) et Embase (Ovid)*, ont été interrogées par la bibliothécaire en utilisant des mots-clés et mots-sujets qui tiennent compte des concepts « intelligence artificielle » et « documentation ».

Les résultats ont été restreints aux publications en anglais et en français, aux cinq dernières années (2020-2025) et aux pays membres de l'OCDE. Les titres et résumés de l'ensemble des références ont été importés dans le logiciel Endnote 20. La liste détaillée des mots-clés et des stratégies de recherche utilisés est fournie à l'Annexe 1. Enfin, les bibliographies des articles retenus ont été examinées manuellement par les deux premiers auteurs (RW, AD) afin d'identifier d'éventuelles références supplémentaires pertinentes.

Sélection des documents

L'ensemble des documents identifiés lors de la recherche documentaire a été examiné conformément aux critères d'inclusion et d'exclusion décrits plus haut (Tableau 1).

La sélection des documents s'est effectuée en deux étapes. D'abord, un premier tri a été réalisé à partir des titres et résumés. Pour assurer la rigueur du processus, 20 % des documents identifiés ont été évalués à l'aveugle par deux auteurs (RW et AD). Le reste a été analysé par un seul auteur (RW). Ensuite, les documents retenus lors de cette première étape ont été examinés en texte intégral. De manière similaire, 20 % des documents ont été évalués à l'aveugle par deux auteurs (RW et AD), tandis que le reste a été

sélectionné par RW. À chaque étape, les désaccords ont été résolus par discussion jusqu'à l'obtention d'un consensus. Le diagramme détaillé du processus de sélection est présenté à l'Annexe 2.

Évaluation de la qualité méthodologique des documents retenus

La qualité méthodologique des études retenues a été évaluée par les deux auteurs principaux (RW et AD) à l'aide du *Mixed-Method Appraisal Tool* (MMAT). Le MMAT est un outil reconnu et validé pour évaluer la qualité méthodologique des études quantitatives, qualitatives et mixtes (Hong et al., 2018). Les résultats de l'analyse de la qualité des études sont présentés à l'Annexe 3.

Extraction et synthèse des données

L'extraction des données a été effectuée par les deux auteurs principaux (RW et AD) à l'aide d'une grille standardisée. Celle-ci comprend notamment les éléments suivants : auteur, année de publication, pays, objectif de l'étude, description de l'intervention, caractéristiques de l'échantillon, devis de recherche, tests statistiques réalisés et principaux résultats.

Une synthèse narrative des données recueillies a ensuite été effectuée. Elle est organisée en deux sections correspondant aux deux questions d'évaluation. Les résultats relatifs aux effets sont présentés selon les différentes variables d'intérêt. La qualité méthodologique des études est ensuite exposée, en soulignant les principales limites à prendre en compte dans l'interprétation des résultats. Enfin, un résumé des principaux constats est présenté à la fin de chaque section sous forme d'encadré intitulé « Sommaire ».

RÉSULTATS

Après élimination des doublons, la recherche documentaire a permis d'identifier 548 références. Toutes les références provenaient des bases de données scientifiques. Aucune nouvelle référence n'a été retrouvée dans les bibliographies consultées. Sur les 548 références initiales, 38 ont été sélectionnées pour une lecture complète. À l'issue du processus de sélection, 20 publications ont satisfait aux critères d'inclusion et ont été retenues pour cette revue rapide. Les détails sur le nombre de documents inclus et exclus à chaque étape sont présentés dans le diagramme de sélection à l'Annexe 2.

La majorité des études proviennent des États-Unis (n= 17). Les autres ont été réalisées au Royaume-Uni (n = 1), Japon (n = 1), et aux Pays-Bas (n=1). La plupart des travaux ont été publiés entre 2023 et 2025, ce qui reflète le caractère très récent de ce champ de recherche.

Les études répertoriées incluent deux essais contrôlés randomisés (ECR) (Baker et al., 2024; Kakaday et al., 2025), deux études pré-post avec groupe témoin (Haberle et al., 2024; Liu et al., 2024), sept études pré-post sans groupe témoin (Albrecht et al., 2025; Duggan et al., 2025; Galloway et al., 2024; Ma et al., 2025; Misurac et al., 2025; Shah, Devon-Sand, et al., 2025; Stults et al., 2025), cinq études intra-sujet (Balloch et al., 2024; Barak-Corren et al., 2024; Omon et al., 2025; van Buchem et al., 2024; Wang et al., 2021), une étude a utilisé une enquête transversale et un devis pré-post sans groupe témoin (Owens et al., 2024), deux études qualitatives (Bundy et al., 2024; Shah, Crowell, et al., 2025) et une étude mixte (Nguyen et al., 2023).

La taille des échantillons variait de 2 à 140. Les échantillons se composaient de cliniciens de première ligne, des médecins spécialistes ainsi que d'étudiants en médecine. Une seule étude s'est intéressée aux professionnels de la réadaptation (Omon et al., 2025) et une autre incluait des ergothérapeutes et des orthophonistes et d'autres spécialités médicales (Balloch et al., 2024). Dans huit études, les participants comptaient plus de dix ans d'expérience clinique, et dans trois autres, entre cinq et dix ans. Les années d'expérience n'étaient pas précisées dans les études restantes.

Quinze études ont évalué l'outil d'IA dans des contextes cliniques réels, où celui-ci était effectivement implanté et utilisé dans la pratique quotidienne des professionnels de santé. Cinq autres études ont testé l'outil dans des conditions expérimentales contrôlées, utilisant par exemple des simulations, des enregistrements audio ou des scénarios standardisés, afin de l'évaluer dans un environnement reproduisant la réalité clinique sans déployer l'outil directement en pratique.

Les études recensées ont eu recours à divers modèles d'intelligence artificielle, incluant des modèles de reconnaissance vocale, des modèles de langage de grande taille et des modèles de traitement automatique du langage naturel. Parmi les outils commerciaux identifiés, DAX Copilot était le plus fréquemment utilisé (n = 10), suivi de ChatGPT (n = 3). Les plateformes *Abridge, Medimo* et *Autoscriber* ont chacune fait l'objet d'une seule étude. Enfin, quatre études ont employé des technologies propriétaires ou développées localement, dont le nom de marque n'était pas précisée.

Les outils d'IA étaient principalement utilisés pour la génération de notes cliniques et de résumés de consultation, et dans une moindre mesure la rédaction de lettres médicales. 11 études ont mentionné que

les participants avaient la possibilité d'éditer les notes. Le reste des études n'ont pas rapporté d'informations à cet effet.

Plus de détails sur les caractéristiques des études et des interventions sont rapportés à l'Annexe 4.

Les résultats des études sont présentés dans la prochaine section qui se divise en deux sections correspondant aux deux guestions évaluées.

I. Impacts sur les professionnels de la santé et leur travail

I.I. Impacts sur les professionnels de la santé

a. Charge de travail liée à la documentation

Dix études ont examiné l'effet de la rédaction assistée par IA sur la charge de travail des cliniciens, en s'appuyant à la fois sur des échelles standardisées et sur des approches qualitatives.

En utilisant le *NASA Task Load Index* (NASA-TLX) et une variante de ce questionnaire, le *Physician Task Load*, deux études pré-post sans groupe témoin ont mis en évidence une réduction significative de la charge de travail (Shah, Devon-Sand, et al., 2025; Stults et al., 2025). Une étude intra-sujet a également montré une diminution significative de la charge perçue sur cinq des six dimensions du NASA-TLX lors de l'utilisation de l'IA (Balloch et al., 2024). Dans une autre étude intra-sujet, une diminution significative de la pression temporelle a été constatée, tandis qu'aucune différence n'a été révélée concernant la charge mentale ou physique, la performance perçue ou l'effort (Omon et al., 2025). Paradoxalement, l'étude a mis en évidence une hausse du niveau de frustration, que les auteurs ont attribuée au faible niveau de littératie numérique chez les cliniciens (Omon et al., 2025).

À l'aide d'échelles de type Likert, une étude pré-post sans groupe témoin a montré que 77 % des cliniciens rapportaient une diminution globale de la charge de documentation (Albrecht et al., 2025). De son côté, une étude intra-sujet a mis en évidence une réduction de l'effort perçu lors de la rédaction d'une note de supervision, mais uniquement pour l'un des quatre cas cliniques évalués (Barak-Corren et al., 2024).

Les approches qualitatives ont montré aussi des effets positifs. Dans une première étude qualitative, plusieurs participants ont rapporté un allègement de la charge cognitive liée aux tâches cliniques (Bundy et al., 2024). Dans une autre étude, 100 % des cliniciens interrogés ont exprimé une perception positive concernant la demande cognitive et 89 % concernant la charge de travail globale (Shah, Crowell, et al., 2025). Dans leurs réponses ouvertes, des participants d'une étude pré-post ont indiqué que si l'utilisation de l'IA ne supprimait pas totalement la charge de la documentation, elle réduisait néanmoins l'effort mental en captant les détails difficiles à mémoriser et en évitant de rédiger une note clinique complète à partir de zéro (Duggan et al., 2025). Enfin, dans le volet qualitatif d'une étude mixte, tous les cliniciens ont indiqué devoir éditer les notes générées avant de les signer, avec une intensité allant de modifications mineures à majeures; toutefois, ils ont signalé que la charge d'édition a diminué progressivement au fil du temps (Nguyen et al., 2023).

b. Épuisement professionnel

Sept études ont évalué les effets des outils d'IA sur l'épuisement professionnel des cliniciens, à l'aide de questionnaires validés ou d'échelles de type Likert.

Deux études pré-post sans groupe témoin, basées sur des échelles de type Likert, ont rapporté des résultats favorables. Dans la première, 99,7 % des répondants déclaraient percevoir une réduction du risque d'épuisement professionnel lié à la documentation après l'implantation de l'IA ambiante (Albrecht et al., 2025). La seconde a montré qu'après l'utilisation de l'IA, les cliniciens ont déclaré percevoir une diminution de leur sentiment d'épuisement (Duggan et al., 2025).

Deux autres études pré-post sans groupe témoin se sont appuyées sur le *Stanford Professional Fulfillment Index* (PFI) et ont également montré une réduction significative de l'épuisement professionnel. La première a évalué uniquement la dimension « épuisement lié au travail » et a rapporté une amélioration significative de ce score (Shah, Devon-Sand, et al., 2025). La deuxième a mis en évidence une amélioration significative pour la dimension « désengagement interpersonnel », sans changement pour les dimensions « épuisement lié au travail » et « épanouissement professionnel » (Misurac et al., 2025). Les analyses par quartiles ont par ailleurs montré que seuls les cliniciens ayant un niveau élevé d'utilisation de l'IA en bénéficiaient (Misurac et al., 2025).

Une autre étude transversale a examiné l'association entre le niveau d'utilisation de l'IA et les scores d'épuisement mesurés à l'aide de l'*Oldenburg Burnout Inventory* (OLBI). Ses résultats ont indiqué qu'un usage intensif était lié à un désengagement moindre, mais aucune association n'a été observée avec le sous-score d'épuisement ni avec le score total du questionnaire (Owens et al., 2024).

Enfin, deux études pré-post sans groupe témoin utilisant le *Mini-Z Burnout Survey* n'ont constaté aucun changement significatif des scores d'épuisement (Nguyen et al., 2023; Stults et al., 2025).

c. Satisfaction professionnelle, bien-être et équilibre travail-vie personnelle

Sept études ont rapporté des résultats en lien avec la satisfaction professionnelle, le bien-être et l'équilibre travail-vie personnelle. Dans l'ensemble, les études ont montré des expériences globalement positives liées à l'utilisation de l'IA.

Une étude pré-post a indiqué que 64 % des cliniciens jugeaient que l'outil d'IA avait accru leur satisfaction (Albrecht et al., 2025). De manière convergente, une autre étude pré-post a indiqué que l'introduction de l'IA a permis de réduire d'environ la moitié la proportion de cliniciens déclarant un impact négatif de la documentation sur leur bien-être (Galloway et al., 2024). Une étude qualitative a confirmé ces observations où la majorité des participants a rapporté une amélioration de leur qualité de vie (Bundy et al., 2024). Enfin, une étude pré-post avec groupe témoin a indiqué que les cliniciens ayant utilisé l'IA obtenaient des scores plus élevés en matière de sécurité et de résilience et décompression, bien qu'elle ne précise pas si ce changement était significatif (Haberle et al., 2024).

Les résultats relatifs à l'équilibre travail-vie personnelle sont également encourageants. Une étude prépost a révélé que 70 % des cliniciens percevaient une amélioration , attribuée à la fois au gain de temps et à une réduction du stress lié à la documentation (Misurac et al., 2025). Une étude qualitative a rapporté

des résultats similaires où 91 % des répondants ont déclaré une meilleure intégration travail-vie personnelle après l'implantation de l'IA (Shah, Crowell, et al., 2025).

En revanche, une étude pré-post avec groupe témoin n'a pas montré d'effets significatifs en indiquant qu'environ la moitié des participants du groupe intervention estimaient que leur expérience du dossier médical électronique (DME) restait inchangée avant et après l'implantation de l'IA (Liu et al., 2024).

d. Engagement des professionnels

Sept études ont examiné l'engagement des professionnels. Dans l'ensemble, les études suggèrent que l'utilisation des outils d'IA améliore la qualité des interactions entre les cliniciens et les patients.

Une étude pré-post a rapporté que 80 % des cliniciens percevaient un impact positif sur leurs interactions avec les patients, attribuable à une meilleure écoute et à une attention accrue, ce qui a renforcé par la même occasion leur satisfaction professionnelle (Misurac et al., 2025). De façon convergente, une autre étude pré-post a révélé que la proportion de cliniciens déclarant pouvoir accorder une attention exclusive aux patients est passée de 57,9 % à 93,0 % après l'implantation de l'IA (Stults et al., 2025). Une autre étude pré-post a montré que l'utilisation de l'IA diminuait significativement la distraction pendant les consultations (Duggan et al., 2025). Enfin, une étude pré-post avec groupe témoin a indiqué que les cliniciens ayant utilisés l'IA obtenaient des scores plus élevés en matière d'engagement, bien que la significativité de ces différences ne soit pas précisée (Haberle et al., 2024).

Une étude intra-sujet a également montré que 100 % des cliniciens estimaient pouvoir accorder toute leur attention aux patients lors de l'utilisation de l'IA, contre 66 % sans l'outil, et que 94 % rapportaient une expérience clinique positive, comparativement à 71 % sans IA (Balloch et al., 2024). Dans le même sens, une étude qualitative a rapporté que la majorité des cliniciens percevaient une interaction plus personnelle et attentive avec leurs patients (Bundy et al., 2024). Des résultats similaires ont été rapportés dans une autre étude qualitative où 68 % des cliniciens exprimaient des perceptions positives concernant leur engagement avec les patients (Shah, Crowell, et al., 2025).

I.II. Impacts sur le travail des professionnels

a. Temps de documentation

Le temps consacré à la documentation a été examiné dans quinze études. Il a été rapporté sous différentes formes : temps par note ou par patient, temps par jour, temps hebdomadaire, ou encore vitesse de rédaction exprimée en mots/minute. Les méthodes de mesure variaient, allant de mesures objectives fondées sur les données des DME ou sur des mesures chronométrées, à des mesures subjectives basées sur la perception déclarée par les cliniciens.

La grande majorité des travaux (11 études) ont rapporté une réduction du temps de documentation ou une amélioration de la vitesse de rédaction.

Quatre études se sont basées sur les données issues des DME. Trois parmi ces études ont employé des devis pré-post sans groupe témoin : une première a révélé une baisse de 20,4 % du temps par rendez-vous (Duggan et al., 2025), une deuxième a montré une réduction moyenne de 1,22 minutes par note et

d'environ 7 minutes par jour (Ma et al., 2025), et une troisième a indiqué une diminution de 28,8 % par note (Owens et al., 2024). Une étude pré-post avec groupe témoin a également confirmé ces résultats en montrant une réduction significative du temps de documentation en faveur de l'outil d'IA par rapport à la rédaction habituelle (Haberle et al., 2024).

Trois études intra-sujet utilisant des mesures chronométrées ont également révélé des résultats positifs : la première étude a montré que la rédaction avec l'outil d'IA est plus rapide que la saisie manuelle (152 s contre 202 s) (van Buchem et al., 2024), la deuxième a rapporté un gain de vitesse allant de 59 à 131 mots/minute en faveur de l'outil d'IA (Wang et al., 2021), et la troisième (N = 10) a observé une réduction de 36 à 43 % du temps pour les cas simples, sans différence pour les cas complexes (Barak-Corren et al., 2024).

Quatre études se sont appuyées sur des mesures auto-rapportées : une étude pré-post avec groupe témoin a indiqué que 43,5 % des cliniciens du groupe intervention déclaraient une réduction du temps, contre 18,2 % dans le groupe contrôle (Liu et al., 2024) ; une étude avant-après sans groupe témoin a rapporté un gain perçu médian de 20 minutes par demi-journée (Shah, Devon-Sand, et al., 2025); une autre étude du même type a montré qu'après implantation, plus de la moitié des cliniciens consacraient une heure ou moins par semaine à la documentation contre 14 % auparavant (Stults et al., 2025) ; enfin, une étude mixte a fait état d'une amélioration du sentiment de disposer de temps suffisant pour la documentation (Nguyen et al., 2023).

En revanche, trois études n'ont pas mis en évidence de changement significatif: un ECR basé sur les données du DME (Kakaday et al., 2025) ainsi qu'un autre ECR (Baker et al., 2024) et une étude intra-sujet utilisant des mesures chronométrées (Omon et al., 2025).

b. Durée des consultations

Deux études se sont intéressées à la durée des consultations. La première, indique que l'utilisation de l'outil d'IA a permis de réduire significativement la durée totale des consultations de 26,3 % sans pour autant changer le temps consacré aux échanges avec les patients et leurs familles (Balloch et al., 2024). Dans la seconde, les participants ont souligné dans leurs réponses ouvertes une diminution de la durée des consultations (Duggan et al., 2025).

c. Temps de travail en dehors des heures régulières

Sept études ont examiné l'impact de la rédaction assistée par IA sur le temps de travail effectué en dehors des heures régulières. Parmi celles s'appuyants sur des mesures extraites du DME, trois études pré-post sans groupe témoin ont rapporté une réduction significative du temps quotidien en dehors des heures régulières de 1,6 minute (Duggan et al., 2025), 4 minutes (Owens et al., 2024), et 7,17 minutes par jour (Ma et al., 2025). En revanche, une étude pré-post avec groupe témoin a observé au contraire une légère augmentation de 4,69 % dans le groupe intervention (Haberle et al., 2024).

Les études basées sur des mesures auto-rapportées vont dans le même sens. Une étude pré-post avec groupe témoin a montré que 47,1 % des cliniciens du groupe intervention déclaraient passer moins de temps sur le DME à domicile, contre 14,5 % dans le groupe contrôle (Liu et al., 2024). Une étude pré-post

sans groupe témoin a indiqué que 71 % des répondants au sondage post intervention estimaient réduire leur temps consacré à la documentation après implantation de l'outil (Albrecht et al., 2025). Enfin, dans une étude qualitative, 9 des 12 participants ont mentionné une réduction significative du travail effectué en dehors des heures régulières (Bundy et al., 2024).

d. Flux de travail

Une étude pré-post sans groupe témoin a rapporté des résultats en lien avec le flux de travail indiquant que les cliniciens ont souligné dans leurs réponses ouvertes une plus grande efficacité dans leur flux de travail ainsi qu'une diminution de la durée des consultations (Duggan et al., 2025). Une autre étude du même type a montré que les cliniciens étaient cinq fois plus susceptibles de pouvoir compléter leurs notes avant la visite suivante après l'implantation de l'IA (Albrecht et al., 2025).

Une description plus détaillée des variables d'intérêt et des outils de mesure est présentée à l'Annexe 5 et des détails sur les résultats et les analyses statistiques utilisées sont rapportés à l'Annexe 6.

Qualités méthodologiques des études

L'interprétation des résultats doit demeurer prudente, compte tenu des diverses limites méthodologiques identifiées dans les études:

- Échantillons non représentatifs : plusieurs études reposaient sur de petits échantillons, des stratégies de recrutement non probabiliste ou des critères d'inclusion insuffisamment précisés.
- Facteurs de confusion non pris en considération : dans une grande majorité des études, certaines variables susceptibles d'influencer les résultats, telles que l'expérience clinique, la discipline professionnelle, la charge clinique habituelle ou le niveau de familiarité avec les technologies numériques, n'ont pas été contrôlées.
- Manque d'informations sur le déroulement des interventions : dans plusieurs études, les auteurs ne précisaient pas les modalités de mise en œuvre de l'intervention ni le degré d'adhésion au protocole.
- Robustesse limitée des devis de recherche : la majorité des travaux utilisaient un devis pré—post sans groupe témoin ou des approches intra-sujets.

Plus de détails sur l'évaluation de la qualité méthodologique des études (MMAT) et les principales limites se trouvent à l'Annexe 3 et à l'Annexe 6.

SOMMAIRE

Impacts des outils d'IA sur les professionnels de la santé et leur travail

Constats généraux

- La majorité des études recensées reposent sur des mesures subjectives.
- Les outils de mesure sont variés, ce qui complique la comparaison directe entre les résultats.
- La **robustesse** des devis est **limitée** : la plupart des travaux sont de type pré—post sans groupe témoin ou intra-sujet.
- La **qualité** de la grande majorité des études est jugée **faible**, en raison notamment de la taille restreinte des échantillons, de leur manque de représentativité et du contrôle insuffisant des facteurs de confusion.

Impacts des outils d'IA sur les professionnels de la santé et leur travail

Impacts des outils d'IA sur les professionnels de la santé

• Charge de travail

- Des effets prometteurs sont rapportés indiquant une réduction perçue de la charge de documentation.
- Les bénéfices pourraient être plus marqués chez les cliniciens disposant d'une meilleure littératie numérique et après une période de familiarisation avec les outils.

Épuisement professionnel

 Les effets sont hétérogènes selon les outils de mesures utilisés et les dimensions évaluées.

• Satisfaction professionnelle, bien-être et équilibre travail-vie personnelle

 Des effets prometteurs sont rapportés, suggérant une meilleure satisfaction professionnelle ainsi qu'un meilleur équilibre travail-vie personnelle, principalement en lien avec le gain de temps et la réduction du stress lié à la documentation.

• Engagement des professionnels

 Les effets sont également jugés prometteurs, avec une amélioration perçue de l'attention et de l'engagement des cliniciens envers leurs patients, ainsi qu'une réduction des distractions durant les consultations.

Impacts des outils d'IA sur le travail des professionnels

• Temps de documentation

- Des effets prometteurs ont été rapportés, indiquant une réduction du temps consacré à la documentation ou une amélioration de la vitesse de rédaction.
- Les gains observés peuvent varier selon la complexité des cas cliniques, les bénéfices étant plus marqués pour les cas simples que pour les cas complexes.

• Durée des consultations

Les données sur la durée des consultations sont insuffisantes.

• Temps de travail en dehors des heures régulières

 Des effets prometteurs ont été rapportés, indiquant une réduction du temps de travail en dehors des heures régulières. Néanmoins, la signification clinique et organisationnelle de cette réduction demeure à clarifier.

Flux de travail

- Les données sur le flux de travail sont insuffisantes.
- → En résumé, l'utilisation d'outils d'IA pour la documentation clinique semble avoir des effets positifs sur les **professionnels de la santé**, en réduisant leur **charge de travail** et en améliorant leur **satisfaction**, leur **équilibre travail-vie personnelle** et leur **engagement**, ainsi que sur **leur travail**, notamment en **diminuant** le **temps consacré à la documentation** et le **travail effectué en dehors des heures régulières**.
- → Bien qu'encourageants, ces résultats doivent être interprétés avec prudence et nécessitent d'être confirmés par des études plus robustes et fondées sur des mesures objectives.

II. Impacts sur les systèmes de santé

Les connaissances disponibles sur les impacts des outils d'IA sur les systèmes de santé sont limitées. Les études recensées se concentrent principalement sur les dimensions d'efficience et de productivité. D'autres dimensions importantes en rapport avec la performance des systèmes, tels que les temps d'attente, la qualité des soins et les coûts associés, demeurent largement inexplorées.

a. Efficience

Deux études se sont intéressées à l'impact de la rédaction assistée par IA sur l'efficience.

Un essai contrôlé randomisé a examiné le rapport entre la qualité et le temps de documentation et a montré que la dictée était significativement plus efficiente que la saisie au clavier, tandis qu'aucune différence n'a été observée entre la dictée et l'outil d'IA (Baker et al., 2024). Par ailleurs, dans une étude intra-sujet basée sur des mesures auto-rapportées, les cliniciens ont jugé les résumés générés par l'outil d'IA comme très efficients, leur attribuant une note moyenne de 8,2/10 sur une échelle de Likert (Barak-Corren et al., 2024).

b. Productivité

Trois études ont rapporté des résultats en lien avec la productivité.

Une étude pré-post sans groupe témoin a montré une augmentation significative de 9,3 % du taux de dossiers fermés le jour même (Duggan et al., 2025). Dans une autre étude pré-post, 58,1 % des cliniciens ont déclaré que l'IA augmentait leur productivité (Galloway et al., 2024). Enfin, une étude pré-post avec groupe témoin a aussi rapporté une hausse statistiquement significative de la productivité mesurée en Work Relative Value Units (wRVU) dans le groupe intervention (Haberle et al., 2024), bien que le changement soit faible (de 1%) et sans signification clinique notable. L'étude a montré également que le nombre moyen de patients attribués à chaque médecin n'a pas changé significativement et, parallèlement, les délais de documentation et de codage se sont accrus dans les deux groupes, sans différence significative attribuable à l'utilisation de l'outil d'IA.

Une description plus détaillée des variables d'intérêt et des outils de mesure est présentée à l'Annexe 5. Pour plus de détails sur les résultats et les analyses statistiques, veuillez consulter l'Annexe 6.

Qualité méthodologique des résultats

Les résultats doivent être interprétés avec précaution, car plusieurs limites méthodologiques ont été relevées dans les études:

- Échantillons peu représentatifs : plusieurs travaux reposaient sur un nombre restreint de participants, des méthodes de recrutement non aléatoires ou des critères d'inclusion mal définis.
- Facteurs de confusion non contrôlés : certaines variables pouvant influencer les résultats telles que l'expérience clinique, la discipline professionnelle, la charge de travail habituelle ou le degré de familiarité avec les technologies numériques n'ont pas été prises en compte.
- Informations incomplètes sur les interventions : dans plusieurs études, les auteurs ne détaillaient pas les modalités d'application de l'intervention ni le niveau de conformité au protocole.
- Devis de recherche avec faible niveau de robustesse : la plupart des études utilisaient un devis pré-post.

Plus de détails sur l'évaluation de la qualité méthodologique des études (MMAT) et les principales limites se trouvent à l'Annexe 3 et à l'Annexe 6.

SOMMAIRE

Impacts des outils d'IA sur les systèmes de santé

Constats généraux

- Plusieurs résultats reposent sur des mesures subjectives.
- Les outils de mesure sont variés, ce qui complique la synthèse et la comparaison directe entre les résultats.
- La robustesse des devis est variable : la plupart des travaux sont de type pré-post.
- La qualité des études est jugée faible à modérée, en raison notamment de la taille restreinte des échantillons, de leur manque de représentativité et du contrôle insuffisant des facteurs de confusion.

Impacts des outils d'IA sur les systèmes de santé

Très peu d'études se sont penchées sur les impacts des outils d'IA au niveau des systèmes de santé. Celles recensées portent principalement sur les dimensions de l'efficience et de la productivité. Aucune étude n'a examiné d'autres dimensions en rapport avec la performance des systèmes tels que les temps d'attente, la qualité des soins ou les coûts associés.

• Efficience

Les données sur l'efficience sont insuffisantes.

Productivité

- Des effets prometteurs ont été rapportés, indiquant une amélioration de la productivité.
- → En résumé, l'utilisation des outiles d'IA pour la documentation clinique semble contribuer à amliorer la productivité des cliniciens. Cependant, leurs impacts sur d'autres dimensions importantes du foctionnement des systèmes de santé demeurent méconnus.
- → Ces résultats doivent être interprétés avec prudence et nécessitent d'être confirmés par des études plus robustes et fondées sur des mesures objectives.

CONCLUSION

Cette revue rapide visait à synthétiser les connaissances actuelles sur les impacts de l'utilisation des outils d'IA pour la documentation clinique sur les professionnels de la santé, leur travail et les systèmes de santé. Vingt (20) études publiées dans les cinq dernières années ont été incluses et analysées.

En ce qui concerne les impacts sur les professionnels de la santé, les résultats indiquent des effets globalement prometteurs sur la réduction de la charge de travail, l'amélioration de la satisfaction professionnelle, de l'équilibre travail-vie personnelle et de l'engagement des professionnels. Cependant, les effets sur l'épuisement professionnel sont variables selon les outils de mesure utilisés et les dimensions évaluées. Pour ce qui est du travail des professionnels, les résultats mettent en évidence des effets prometteurs, notamment une réduction du temps consacré à la documentation et du travail effectué en dehors des heures régulières. En revanche, en raison du manque de données, les résultats demeurent non concluants quant au temps total des consultations et au flux de travail. Enfin, très peu d'études ont examiné les impacts des outils d'IA sur les systèmes de santé; leurs résultats suggèrent des effets positifs sur la productivité, mais demeurent non concluants en ce qui concerne l'efficience en raison du manque de données. Aucune étude n'a examiné d'autres aspects de la performance des systèmes tels que les temps d'attente, la qualité des soins ou les coûts associés.

Ces résultats doivent être interprétés avec prudence en raison des limites méthodologiques importantes des études recensées en lien notamment avec la représentativité des échantillons et du contrôle insuffisant des facteurs de confusions, de la nature subjective des mesures utilisées, de la variabilité des contextes d'implantation et du manque de données probantes à long terme. Des recherches supplémentaires rigoureuses sont nécessaires pour confirmer ces effets et développer des connaissances permettant d'améliorer la généralisation des résultats.

Forces et limites de la revue rapide

Cette revue rapide dresse un portrait actualisé des publications récentes portant sur l'utilisation des outils d'IA dans la documentation clinique. Elle s'appuie sur une analyse méthodique de la littérature scientifique, et les constats présentés reposent sur une synthèse rigoureuse et critique des données disponibles. Dans l'ensemble, cette revue constitue un appui pertinent à la prise de décision concernant l'adoption des outils d'IA dans les services de réadaptation, en mettant en évidence leurs effets potentiels sur la pratique clinique.

Cependant, certaines limites doivent être reconnues. D'une part, cette revue n'avait pas pour objectif d'être exhaustive, ce qui signifie que certaines études pertinentes ont pu être omises. D'autre part, elle n'aborde pas certaines dimensions essentielles, telles que la dimension coût-efficacité des outils d'IA, la qualité de la documentation produite, ainsi que la sécurité et la confidentialité des données.

RÉFÉRENCES

- Albrecht, M., Shanks, D., Shah, T., Hudson, T., Thompson, J., Filardi, T., Wright, K., Ator, G. A., & Smith, T. R. (2025). Enhancing clinical documentation with ambient artificial intelligence: A quality improvement survey assessing clinician perspectives on work burden, burnout, and job satisfaction. *JAMIA Open*, 8(1), ooaf013. https://doi.org/10.1093/jamiaopen/ooaf013
- Baker, H. P., Dwyer, E., Kalidoss, S., Hynes, K., Wolf, J., & Strelzow, J. A. (2024). ChatGPT's Ability to Assist with Clinical Documentation: A Randomized Controlled Trial. *The Journal of the American Academy of Orthopaedic Surgeons*, 32(3), 123-129. https://doi.org/10.5435/JAAOS-D-23-00474
- Balloch, J., Sridharan, S., Oldham, G., Wray, J., Gough, P., Robinson, R., Sebire, N. J., Khalil, S., Asgari, E., Tan, C., Taylor, A., & Pimenta, D. (2024). Use of an ambient artificial intelligence tool to improve quality of clinical documentation. *Future Healthcare Journal*, 11(3), 100157. https://doi.org/10.1016/j.fhj.2024.100157
- Barak-Corren, Y., Wolf, R., Rozenblum, R., Creedon, J. K., Lipsett, S. C., Lyons, T. W., Michelson, K. A., Miller, K. A., Shapiro, D. J., Reis, B. Y., & Fine, A. M. (2024). Harnessing the Power of Generative AI for Clinical Summaries: Perspectives From Emergency Physicians. *Annals of Emergency Medicine*, 84(2), 128-138. https://doi.org/10.1016/j.annemergmed.2024.01.039
- Bongurala, A. R., Save, D., Virmani, A., & Kashyap, R. (2024). Transforming Health Care With Artificial Intelligence: Redefining Medical Documentation. *Mayo Clinic Proceedings: Digital Health*, *2*(3), 342-347. https://doi.org/10.1016/j.mcpdig.2024.05.006
- Bracken, A., Reilly, C., Feeley, A., Sheehan, E., Merghani, K., & Feeley, I. (2025). Artificial Intelligence (AI) Powered Documentation Systems in Healthcare: A Systematic Review. *Journal of Medical Systems*, 49(1), 28. https://doi.org/10.1007/s10916-025-02157-4
- Bundy, H., Gerhart, J., Baek, S., Connor, C. D., Isreal, M., Dharod, A., Stephens, C., Liu, T.-L., Hetherington, T., & Cleveland, J. (2024). Can the Administrative Loads of Physicians be Alleviated by Al-Facilitated Clinical Documentation? *Journal of General Internal Medicine*, 39(15), 2995-3000. https://doi.org/10.1007/s11606-024-08870-z
- Duggan, M. J., Gervase, J., Schoenbaum, A., Hanson, W., Howell, J. T., Sheinberg, M., & Johnson, K. B. (2025). Clinician Experiences With Ambient Scribe Technology to Assist With Documentation Burden and Efficiency. *JAMA Network Open*, 8(2), e2460637. https://doi.org/10.1001/jamanetworkopen.2024.60637
- Gaffney, A., Woolhandler, S., Cai, C., Bor, D., Himmelstein, J., McCormick, D., & Himmelstein, D. U. (2022). Medical Documentation Burden Among US Office-Based Physicians in 2019: A National Study. *JAMA Internal Medicine*, 182(5), 564-566. https://doi.org/10.1001/jamainternmed.2022.0372

- Galloway, J. L., Munroe, D., Vohra-Khullar, P. D., Holland, C., Solis, M. A., Moore, M. A., & Dbouk, R. H. (2024). Impact of an Artificial Intelligence-Based Solution on Clinicians' Clinical Documentation Experience: Initial Findings Using Ambient Listening Technology. *Journal of General Internal Medicine*, 39(13), 2625-2627. https://doi.org/10.1007/s11606-024-08924-2
- Garritty, C., Gartlehner, G., Nussbaumer-Streit, B., King, V. J., Hamel, C., Kamel, C., Affengruber, L., & Stevens, A. (2021). Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. Journal of Clinical Epidemiology, 130, 13-22. https://doi.org/10.1016/j.jclinepi.2020.10.007
- Haberle, T., Cleveland, C., Snow, G. L., Barber, C., Stookey, N., Thornock, C., Younger, L., Mullahkhel, B., & Ize-Ludlow, D. (2024). The impact of nuance DAX ambient listening AI documentation: A cohort study. *Journal of the American Medical Informatics Association: JAMIA*, 31(4), 975-979. https://doi.org/10.1093/jamia/ocae022
- Hamel, C., Michaud, A., Thuku, M., Affengruber, L., Skidmore, B., Nussbaumer-Streit, B., Stevens, A., & Garritty, C. (2020). Few evaluative studies exist examining rapid review methodology across stages of conduct: A systematic scoping review. Journal of Clinical Epidemiology, 126, 131-140. https://doi.org/10.1016/j.jclinepi.2020.06.027
- Hamel, C., Michaud, A., Thuku, M., Skidmore, B., Stevens, A., Nussbaumer-Streit, B., & Garritty, C. (2021). Defining Rapid Reviews: A systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. Journal of Clinical Epidemiology, 129, 74-85. https://doi.org/10.1016/j.jclinepi.2020.09.041
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C. et Vedel, I. (2018 Mixed Methods Appraisal Tool (MMAT), version 2018. Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada. article. http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127425851/MMAT_2018_c riteria-manual_2018-04-04.pd
- Kakaday, R., Herrera, E. Z., Coskey, O., Hertel, A. W., & Kaiser, P. (2025). The STREAMLINE Pilot—Study on Time Reduction and Efficiency in Al-Mediated Logging for Improved Note-Taking Experience. *Applied Clinical Informatics*. https://doi.org/10.1055/a-2559-5791
- Le Collège des médecins de famille du Canada. (2015). Guide de conseils pratiques : Mise en place d'un système de DMÉ. *Patient's Medical Home*. https://patientsmedicalhome.ca/fr/resources/conseils-pratiques/guide-de-conseils-pratiques-mise-en-place-dun-systeme-de-dme/
- Liu, T.-L., Hetherington, T. C., Stephens, C., McWilliams, A., Dharod, A., Carroll, T., & Cleveland, J. A. (2024).

 Al-Powered Clinical Documentation and Clinicians' Electronic Health Record Experience: A

 Nonrandomized Clinical Trial. *JAMA Network Open*, 7(9), e2432460.

 https://doi.org/10.1001/jamanetworkopen.2024.32460

- Ma, S. P., Liang, A. S., Shah, S. J., Smith, M., Jeong, Y., Devon-Sand, A., Crowell, T., Delahaie, C., Hsia, C., Lin, S., Shanafelt, T., Pfeffer, M. A., Sharp, C., & Garcia, P. (2025). Ambient artificial intelligence scribes: Utilization and impact on documentation time. *Journal of the American Medical Informatics Association: JAMIA*, 32(2), 381-385. https://doi.org/10.1093/jamia/ocae304
- Misurac, J., Knake, L. A., & Blum, J. M. (2025). The Effect of Ambient Artificial Intelligence Notes on Provider Burnout. *Applied Clinical Informatics*, *16*(2), 252-258. https://doi.org/10.1055/a-2461-4576
- Muhiyaddin, R., Elfadl, A., Mohamed, E., Shah, Z., Alam, T., Abd-Alrazaq, A., & Househ, M. (2022). Electronic Health Records and Physician Burnout: A Scoping Review. *Studies in Health Technology and Informatics*, 289, 481-484. https://doi.org/10.3233/SHTI210962
- Nguyen, O. T., Turner, K., Charles, D., Sprow, O., Perkins, R., Hong, Y.-R., Islam, J. Y., Khanna, N., Alishahi Tabriz, A., Hallanger-Johnson, J., Bickel Young, J., & Moore, C. E. (2023). Implementing Digital Scribes to Reduce Electronic Health Record Documentation Burden Among Cancer Care Clinicians:

 A Mixed-Methods Pilot Study. *JCO Clinical Cancer Informatics*, 7, e2200166. https://doi.org/10.1200/CCI.22.00166
- Omon, K., Sasaki, T., Koshiro, R., Fuchigami, T., & Hamashima, M. (2025). Effects of Introducing Generative AI in Rehabilitation Clinical Documentation. *Cureus*, *17*(3), e81313. https://doi.org/10.7759/cureus.81313
- Owens, L. M., Wilda, J. J., Hahn, P. Y., Koehler, T., & Fletcher, J. J. (2024). The association between use of ambient voice technology documentation during primary care patient encounters, documentation burden, and provider burnout. *Family Practice*, *41*(2), 86-91. https://doi.org/10.1093/fampra/cmad092
- Shah, S. J., Crowell, T., Jeong, Y., Devon-Sand, A., Smith, M., Yang, B., Ma, S. P., Liang, A. S., Delahaie, C., Hsia, C., Shanafelt, T., Pfeffer, M. A., Sharp, C., Lin, S., & Garcia, P. (2025). Physician Perspectives on Ambient Al Scribes. *JAMA Network Open*, 8(3), e251904. https://doi.org/10.1001/jamanetworkopen.2025.1904
- Shah, S. J., Devon-Sand, A., Ma, S. P., Jeong, Y., Crowell, T., Smith, M., Liang, A. S., Delahaie, C., Hsia, C., Shanafelt, T., Pfeffer, M. A., Sharp, C., Lin, S., & Garcia, P. (2025). Ambient artificial intelligence scribes: Physician burnout and perspectives on usability and documentation burden. *Journal of the American Medical Informatics Association: JAMIA*, 32(2), 375-380. https://doi.org/10.1093/jamia/ocae295
- Stults, C. D., Deng, S., Martinez, M. C., Wilcox, J., Szwerinski, N., Chen, K. H., Driscoll, S., Washburn, J., & Jones, V. G. (2025). Evaluation of an Ambient Artificial Intelligence Documentation Platform for Clinicians. *JAMA Network Open*, 8(5), e258614. https://doi.org/10.1001/jamanetworkopen.2025.8614
- Tricco, A. C., Cardoso, R., Thomas, S. M., Motiwala, S., Sullivan, S., Kealey, M. R., Hemmelgarn, B., Ouimet, M., Hillmer, M. P., Perrier, L., Shepperd, S., & Straus, S. E. (2016). Barriers and facilitators to uptake

- of systematic reviews by policy makers and health care managers: A scoping review. Implementation Science, 11(1), 4. https://doi.org/10.1186/s13012-016-0370-1
- Tricco, A. C., Garritty, C. M., Boulos, L., Lockwood, C., Wilson, M., McGowan, J., McCaul, M., Hutton, B., Clement, F., Mittmann, N., Devane, D., Langlois, E. V., Abou-Setta, A. M., Houghton, C., Glenton, C., Kelly, S. E., Welch, V. A., LeBlanc, A., Wells, G. A., ... Straus, S. E. (2020). Rapid review methods more challenging during COVID-19: Commentary with a focus on 8 knowledge synthesis steps. Journal of Clinical Epidemiology, 126, 177. https://doi.org/10.1016/j.jclinepi.2020.06.029
- van Buchem, M. M., Kant, I. M. J., King, L., Kazmaier, J., Steyerberg, E. W., & Bauer, M. P. (2024). Impact of a Digital Scribe System on Clinical Documentation Time and Quality: Usability Study. *Jmir Ai*, *3*, e60020. https://doi.org/10.2196/60020
- Wang, J., Lavender, M., Hoque, E., Brophy, P., & Kautz, H. (2021). A patient-centered digital scribe for automatic medical documentation. *JAMIA Open*, *4*(1), ooab003. https://doi.org/10.1093/jamiaopen/ooab003

ANNEXE 1. RECHERCHE DOCUMENTAIRE

Bases de données scientifiques consultées

Base de données	Platerforme	Couverture	Date de consultation	Nombre de références
1. Medline	Ovid	Ovid MEDLINE(R) ALL 1946 to May 06, 2025	Mai 2025	281
2. All EMB Reviews	Ovid	 EBM Reviews - Cochrane Database of Systematic Reviews 2005 to April 28, 2025 EBM Reviews - ACP Journal Club 1991 to April 2025 EBM Reviews - Database of Abstracts of Reviews of Effects 1st Quarter 2016 EBM Reviews - Cochrane Clinical Answers April 2025 EBM Reviews - Cochrane Central Register of Controlled Trials March 2025 EBM Reviews - Cochrane Methodology Register 3rd Quarter 2012 EBM Reviews - Health Technology Assessment 4th Quarter 2016 EBM Reviews - NHS Economic Evaluation Database 1st Quarter 2016 	Mai 2025	29
3. Embase	Ovid	Embase 1980 to 2025 Week 18	Mai 2025	451
	1		Total des bases	761
			TOTAL (sans les doublons)	548

Stratégies de recherche dans les bases de données

Medline

#	Concept	Stratégie de recherche	Résultats
1	Artificial Intelligence	Artificial Intelligence/ OR Ambient Intelligence/ OR (artificial intelligence OR AI OR ambient voice technolog* OR (ambient AJD3 intelligence) OR (ambient ADJ3 AI)).ti,kf,kw,jw OR (artificial intelligence OR AI OR ambient voice technolog* OR (ambient AJD3 intelligence) OR (ambient ADJ3 AI)).ab/freq=2 OR Machine Intelligence.jn	122 032
2	Documentation	Documentation/ OR (documentation? OR scribe? OR scribing).ti,kf,kw OR (documentation? OR scribe? OR scribing).ab/freq=2	32 557
3		(digital scribe? OR digital scribing OR medical scribe? OR medical scribing OR clinical scribe? OR clinical scribing OR ((automatic OR automated OR automating) ADJ3 documentation?) OR (ambient ADJ5 documentation? ADJ5 technolog*) OR ambient scribe? technolog* OR ambient scribing technolog*).ti,kf,kw OR (digital scribe? OR digital scribing OR medical scribe? OR medical scribing OR clinical scribe? OR clinical scribing OR ((automatic OR automated OR automating) ADJ3 documentation?) OR (ambient ADJ5 documentation? ADJ5 technolog*) OR ambient scribe? technolog* OR ambient scribing technolog*).ab/freq=2	215
4	Filtre pays OCDE	afghanistan/ or africa/ or africa, northern/ or africa, central/ or africa, eastern/ or "africa south of the sahara"/ or africa, southern/ or africa, western/ or albania/ or algeria/ or andorra/ or angola/ or "antigua and barbuda"/ or argentina/ or armenia/ or azerbaijan/ or bahamas/ or bahrain/ or bangladesh/ or barbados/ or belize/ or benin/ or bhutan/ or bolivia/ or borneo/ or "bosnia and herzegovina"/ or botswana/ or brazil/ or brunei/ or bulgaria/ or burkina faso/ or burundi/ or cabo verde/ or cambodia/ or cameroon/ or central african republic/ or chad/ or exp china/ or comoros/ or congo/ or "cote d'ivoire"/ or croatia/ or cuba/ or "democratic republic of the congo"/ or cyprus/ or djibouti/ or dominica/ or dominican republic/ or ecuador/ or egypt/ or el salvador/ or equatorial guinea/ or eritrea/ or eswatini/ or ethiopia/ or fiji/ or gabon/ or gambia/ or "georgia (republic)"/ or ghana/ or grenada/ or guatemala/ or guinea/ or guinea-bissau/ or guyana/ or haiti/ or honduras/ or independent state of samoa/ or exp india/ or indian ocean islands/ or indochina/ or indonesia/ or iran/ or iraq/ or jamaica/ or jordan/ or kazakhstan/ or kenya/ or kosovo/ or kuwait/ or kyrgyzstan/ or laos/ or lebanon/ or liechtenstein/ or lesotho/ or liberia/ or libya/ or madagascar/ or malaysia/ or malawi/ or mali/ or malta/ or mauritians/ or montenegro/ or morocco/ or mozambique/ or myanmar/ or namibia/ or nepal/ or nicaragua/ or niger/ or nigeria/ or oman/ or pakistan/ or palau/ or exp panama/ or papua new guinea/ or paraguay/ or peru/ or philippines/ or qatar/ or "republic of belarus"/ or "republic of north macedonia"/ or romania/ or exp russia/ or rwanda/ or "saint kitts and nevis"/ or saint lucia/ or "saint vincent and the grenadines"/ or "sao tome and principe"/ or saudi arabia/ or serbia/ or sierra leone/ or senegal/ or seychelles/ or singapore/ or somalia/ or south africa/ or south sudan/ or sri lanka/ or sudan/ or suriname/ or syria/ or taiwan/ or tajikistan/ or tanzania/ or thailand/ or timor-leste/ or togo/ or tonga/ or "	1 431 393
5	Filtre pays OCDE	"Organisation for Economic Co-Operation and Development"/	686
6	Filtre pays OCDE	australasia/ or exp australia/ or austria/ or baltic states/ or belgium/ or exp canada/ or chile/ or colombia/ or costa rica/ or czech republic/ or exp denmark/ or estonia/ or europe/ or finland/ or exp france/ or exp germany/ or greece/ or hungary/ or iceland/ or ireland/ or israel/ or exp italy/ or exp japan/ or korea/ or latvia/ or lithuania/ or luxembourg/ or mexico/ or netherlands/ or new zealand/ or north america/ or exp norway/ or poland/ or portugal/ or exp "republic of korea"/ or "scandinavian and nordic countries"/ or slovakia/ or slovenia/ or spain/ or sweden/ or switzerland/ or turkey/ or exp united kingdom/ or exp united states/	3 675 839
7	Filtre pays OCDE	European Union/	18 516
8	Filtre pays OCDE	Developed Countries/	21 816

9	Filtre pays OCDE	OR/5-8	3 692 726
10	Filtre pays OCDE	4 NOT 9	1 337 949
11	Combinaison	(((1 AND 2) OR 3) NOT 10) AND (english or french).lg	648
12	Limitations	l/ 11 yr=2020-2025	281

All EBM Reviews

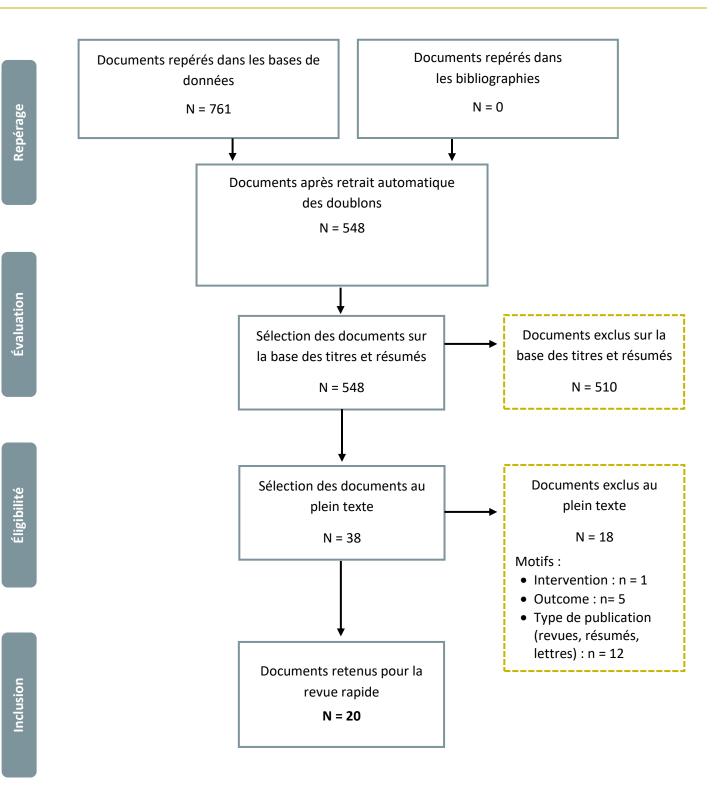
#	Concept	Stratégie de recherche	Résultats
1	Artificial Intelligence	Artificial Intelligence/ OR Ambient Intelligence/ OR (artificial intelligence OR AI OR ambient voice technolog* OR (ambient AJD3 intelligence) OR (ambient ADJ3 AI)).ti,kf,kw,jw OR (artificial intelligence OR AI OR ambient voice technolog* OR (ambient AJD3 intelligence) OR (ambient ADJ3 AI)).ab/freq=2 OR Machine Intelligence.jn	4 994
2	Documentation	Documentation/ OR (documentation? OR scribe? OR scribing).ti,kf,kw OR (documentation? OR scribe? OR scribing).ab/freq=2	3 076
3		(digital scribe? OR digital scribing OR medical scribe? OR medical scribing OR clinical scribe? OR clinical scribing OR ((automatic OR automated OR automating) ADJ3 documentation?) OR (ambient ADJ5 documentation? ADJ5 technolog*) OR ambient scribe? technolog* OR ambient scribing technolog*).ti,kf,kw OR (digital scribe? OR digital scribing OR medical scribe? OR medical scribing OR clinical scribe? OR clinical scribing OR ((automatic OR automated OR automating) ADJ3 documentation?) OR (ambient ADJ5 documentation? ADJ5 technolog*) OR ambient scribe? technolog* OR ambient scribing technolog*).ab/freq=2	18
4	Combinaison	((1 AND 2) OR 3) AND (english or french).lg	41
5	Limitations	l/ 4 yr=2020-2025	29

Embase

#	Concept	Stratégie de recherche	Résultats
1	Artificial Intelligence	exp artificial intelligence/ OR (artificial intelligence OR AI OR ambient voice technolog* OR (ambient AJD3 intelligence) OR (ambient ADJ3 AI)).ti,kw,jx OR (artificial intelligence OR AI OR ambient voice technolog* OR (ambient AJD3 intelligence) OR (ambient ADJ3 AI)).ab/freq=2	182 155
2	Documentation	documentation/ OR (documentation? OR scribe? OR scribing).ti,kw OR (documentation? OR scribe? OR scribing).ab/freq=2	75 255
3		(digital scribe? OR digital scribing OR medical scribe? OR medical scribing OR clinical scribe? OR clinical scribing OR ((automatic OR automated OR automating) ADJ3 documentation?) OR (ambient ADJ5 documentation? ADJ5 technolog*) OR ambient scribe? technolog* OR ambient scribing technolog*).ti,kw OR (digital scribe? OR digital scribing OR medical scribe? OR medical scribing OR clinical scribe? OR clinical scribing OR ((automatic OR automated OR automating) ADJ3 documentation?) OR (ambient ADJ5 documentation? ADJ5 technolog*) OR ambient scribe? technolog* OR ambient scribing technolog*).ab/freq=2	245
4	Filtre pays OCDE	afghanistan/ or africa/ or "africa south of the sahara"/ or albania/ or algeria/ or andorra/ or angola/ or argentina/ or "antigua and barbuda"/ or armenia/ or exp azerbaijan/ or bahamas/ or bahrain/ or bangladesh/ or barbados/ or belarus/ or belize/ or benin/ or bhutan/ or bolivia/ or borneo/ or exp "bosnia and herzegovina"/ or botswana/ or exp brazil/ or brunei darussalam/ or bulgaria/ or burkina faso/ or burundi/ or cambodia/ or cameroon/ or cape verde/ or central africa/ or central african republic/ or chad/ or exp china/ or comoros/ or congo/ or cook islands/ or cote d'ivoire/ or croatia/ or cuba/ or cyprus/ or democratic republic congo/ or djibouti/ or dominica/ or dominican republic/ or ecuador/ or el salvador/ or egypt/ or equatorial guinea/ or eritrea/ or eswatini/ or ethiopia/ or exp "federated states of micronesia"/ or fiji/ or gabon/ or gambia/ or exp "georgia (republic)"/ or ghana/ or grenada/ or guatemala/ or guinea/ or guinea-bissau/ or guyana/ or haiti/ or honduras/ or exp india/ or exp indonesia/ or iran/ or exp irad/ or jamaica/ or jordan/ or kazakhstan/ or kenya/ or kiribati/ or kosovo/ or kuwait/ or kyrgyzstan/ or laos/ or lebanon/ or liechtenstein/ or lesotho/ or liberia/ or libyan arab jamahiriya/ or madagascar/ or malawi/ or exp malaysia/ or maldives/ or mali/ or malta/ or mauritania/ or mauritius/ or melanesia/ or moldova/ or monaco/ or mongolia/ or "montenegro (republic)"/ or morocco/ or mozambique/ or myanmar/ or namibia/ or nauru/ or nepal/ or nicaragua/ or niger/ or nigeria/ or niue/ or north africa/ or oman/ or exp pakistan/ or palau/ or palestine/ or panama/ or papua new guinea/ or paraguay/ or peru/ or philippines/ or polynesia/ or saudi arabia/ or senegal/ or exp serbia/ or seychelles/ or sierra leone/ or singapore/ or "sao tome and principe"/ or solomon islands/ or exp somalia/ or south africa/ or south asia/ or south sudan/ or exp southeast asia/ or sri lanka/ or sudan/ or suriname/ or syrian arab republic/ or taiwan/ or tajikistan/ or tanzania/ or turkmenistan/ or tuvalu/ or ug	1 836 521
5	Filtre pays OCDE	"Organisation for Economic Co-Operation and Development"/	3 309
6	Filtre pays OCDE	exp australia/ or "australia and new zealand"/ or austria/ or baltic states/ or exp belgium/ or exp canada/ or chile/ or colombia/ or costa rica/ or czech republic/ or denmark/ or estonia/ or europe/ or exp finland/ or exp france/ or exp germany/ or greece/ or hungary/ or iceland/ or ireland/ or israel/ or exp italy/ or japan/ or korea/ or latvia/ or lithuania/ or luxembourg/ or exp mexico/ or netherlands/ or new zealand/ or north america/ or exp norway/ or poland/ or exp portugal/ or scandinavia/ or sweden/ or slovakia/ or slovenia/ or south korea/ or exp spain/ or switzerland/ or "turkey (republic)"/ or exp united kingdom/ or exp united states/ or western europe/	3 848 901
7	Filtre pays OCDE	european union/	33 000
8	Filtre pays OCDE	developed country/	34 808

Filtre pays OCDE	OR/5-8	3 884 358
Filtre pays OCDE	4 NOT 9	1 673 722
Combinaison	(((1 AND 2) OR 3) NOT 10) AND (english or french).lg	967
Limitations	l/ 11 yr=2020-2025	451
	Filtre pays OCDE	Filtre pays OCDE 4 NOT 9 Combinaison (((1 AND 2) OR 3) NOT 10) AND (english or french).lg

ANNEXE 2. DIAGRAMME DE SÉLECTION



Source: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

ANNEXE 3. ÉVALUATION DE LA QUALITÉ DES ÉTUDES (MMAT)

Choix de réponse possibles : Oui, Non, Ne sait pas (NSP)

Étude quantitative à répartition aléatoire

Critères	Baker (2024)	Kakaday (2025)
La répartition au hasard des participants (randomisation) est-elle effectuée de manière appropriée ?	NSP	NSP
Les groupes sont-ils comparables au début de l'étude (avant l'intervention) ?	NSP	Non
Les données sur les effets (<i>outcomes</i>) sont-elles complètes ?	Oui	Oui
Est-ce que l'évaluation est effectuée à l'aveugle (les évaluateurs ne savent pas qui reçoit quel traitement) ?	Oui	Non
Les participants ont-ils reçu l'intervention qui leur a été assignée ?	Oui	Oui

Études quantitatives sans répartition aléatoire

Critères	Balloch (2024)	Barak- Corren (2024)	Duggan (2025)	Galloway (2024)	Harberle (2024)	Liu (2024)	Ma (2025)
Les participants constituent-ils un échantillon représentatif de la population cible ?	Non	Non	Non	Non	Oui	Non	Non
Les mesures sont-elles appropriées en ce qui a trait aux effets (outcomes) et à l'intervention (ou l'exposition) ?	Oui	Oui	Oui	Non	Oui	Non	Oui
Les données sur les effets (outcomes) sont-elles complètes ?	NSP	Oui	Oui	Non	Non	NSP	Oui
Les facteurs de confusion sont-ils pris en compte dans la conception de l'étude et l'analyse des données ?	Non	Oui	Oui	Non	Oui	Non	Non
Pendant la période de l'étude, est-ce que l'intervention a été menée (ou l'exposition a eu lieu) comme prévu ?	Oui	Oui	NSP	Non	NSP	NSP	NSP

Critères	Misurak (2025)	Nguyen (2023)	Omon (2025)	Owens (2024)	Shah, Devon- Sand, et al., (2025)	Stults 2025	van Buchem (2024)	Wang (2021)
Les participants constituent-ils un échantillon représentatif de la population cible ?	Non	Non	Oui	Non	Non	Non	Non	Non
Les mesures sont-elles appropriées en ce qui a trait aux effets (outcomes) et à l'intervention (ou l'exposition)?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Les données sur les effets (outcomes) sont-elles complètes ?	Oui	Non	Oui	Non	Oui	Non	Oui	Oui
Les facteurs de confusion sont-ils pris en compte dans la conception de l'étude et l'analyse des données ?	Non	Non	Non	Oui	Non	Non	Non	Non
Pendant la période de l'étude, est-ce que l'intervention a été menée (ou l'exposition a eu lieu) comme prévu ?	NSP	Oui	Oui	Non	Oui	Non	Oui	Oui

Études quantitatives descriptives

Critères	Albercht (2025)
La stratégie d'échantillonnage est-elle pertinente pour répondre à la question de recherche ?	Non
L'échantillon est-il représentatif de la population cible ?	Non
Les mesures sont-elles appropriées ?	Non
Le risque de biais de non-réponse est-il faible ?	NSP
L'analyse statistique est-elle appropriée pour répondre à la question de recherche ?	Non

Études qualitatives

Critères	Bundy (2024)	Shah, Crowell, et al., (2025)
L'approche qualitative est-elle	Oui	Oui
appropriée pour répondre à la question de recherche ?	Oui	Oui
Les méthodes de collecte de données	O i	0
qualitatives sont-elles adéquates pour répondre à la question de recherche ?	Oui	Oui
Les résultats émanent-ils adéquatement des données ?	Oui	Oui
L'interprétation des résultats est-elle suffisamment étayée par les données ?	Oui	Oui
Y a-t-il une cohérence entre les sources, la collecte, l'analyse et l'interprétation des données qualitatives ?	Oui	Oui

ANNEXE 4. CARACTÉRISTIQUES DES ÉTUDES

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
Albercht 2025 États-Unis	Étude descriptive avec une enquête pré-post N.B. Le groupe de répondants au prétest n'était pas exactement composé des mêmes individus que celui ayant répondu au posttest, et les questionnaires administrés présentaient certaines différences entre les deux temps de mesure. Temps de mesure T0: avant et T1: après utilisation de l'IA Le temps entre T0 et T1 était de 92 jours en médiane	Intervention Abridge: une plateforme d'IA ambiante qui résume les conversations médicales pour les cliniciens et les patients dans divers contextes de soins. Comparateur/ Condition (s) de comparaison N/A	93 personnes ont complété le sondage pré- implantation et 99 ont complété le sondage post-implantation Âge (médiane) : 41 ans (37–46) 48 % hommes Spécialités : 34 % en médecine de première ligne, 41 % en surspécialité médicale et 24 % en surspécialité chirurgicale. Années d'expérience : non rapportées
Baker 2024 États-Unis	Devis ECR (randomisation pourtrois techniques de rédaction différentes). Temps de mesure Un seul temps de mesure	Intervention ChatGPT-4 a été utilisé pour rédiger l'histoire de la maladie actuelle à partir d'entretiens standardisés, en suivant une commande et un modèle uniformisés : « Veuillez rédiger un historique de la	N= 11 (33 histoires de maladie ont été examinées) Âge: non rapporté Proportions hommes/ femmes: non rapportées

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
		maladie actuelle du patient en incluant les points suivants : ». Comparateur/ Condition (s) de comparaison Deux autres techniques de documentation ont été utilisées : dictée et saisie au clavier (pas de détails sur ces deux techniques) Les participants devaient interviewer trois patients standardisés (patients-acteurs) et documenter l'histoire de la maladie de chaque patient en utilisant la méthode qui leur avait été attribuée.	Spécialité: 4 étudiants en médecine, 4 résidents en chirurgie orthopédique et 3 chirurgiens titulaires Années d'expérience: non rapportées
Balloch 2024 Royaume-Uni	Devis Étude intra-sujet Chaque clinicien a réalisé trois consultations simulées avec le dossier médical électronique seul (condition contrôle) et trois consultations avec l'assistance de l'outil d'IA (condition intervention). Temps de mesure N/A	Intervention ChatGPT-4 a été utilisé pour générer une note clinique et une lettre de suivi à partir de consultations simulées. Comparateur/ Condition (s) de comparaison Pratique habituelle de documentation (sans IA)	Age: non rapporté Proportions hommes/ femmes: non rapportées Spécialités: Les spécialités couvertes comprenaient l'ergothérapie, l'oncologie, la cardiologie, l'orthophonie, la pneumologie, l'urologie et la dermatologie. Années d'expérience: non rapportées

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
Barak- Corrren 2024 États-Unis	Etude intra-sujet imbriquée dans une étude mixte N.B. Cette revue rapide s'intéresse et rapporte les résultats du volet quantitatif. Chaque participant devait rédiger une note de supervision sans l'utilisation de l'IA pour deux cas cliniques et une note de supervision à partir d'un résumé généré par ChatGPT pour deux autres cas. Temps de mesure N/A	Intervention ChatGPT-4 Plus a été utilisé pour générer une note de supervision médicale destinée à un médecin titulaire à partir des notes rédigées par des résidents pour quatre cas cliniques (deux simples et deux complexes). Comparateur/ Condition (s) de comparaison Pratique habituelle de documentation (sans IA)	N=10 Âge: 60 % avaient moins de 50 ans 50 % hommes Spécialités: médecins urgentistes pédiatriques Années d'expérience: 90 % comptaient plus de 4 ans de pratique
Bundy 2024 États-Unis	Devis Étude qualitative Un seul temps de mesure	Intervention DAX Copilot est un outil qui utilise l'écoute ambiante et des modèles de langage de grande taille pour créer la documentation clinique à partir des consultations. Dans cette étude, DAX a été utilisé pour générer automatiquement des résumés narratifs des conversations clinicien- patient.	 N = 12 Âge : non rapporté 66 % hommes Spécialités : médecins de soins primaires (médecine familiale, médecine interne, médecine pédiatrique) Années d'expérience : non rapportées

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
		Comparateur/ Condition (s) de comparaison Aucun	
Duggan 2025 États-Unis	Devis Étude pré-post sans groupe témoin Temps de mesure T0: avant; T1: 7 semaines après l'utilisation de l'IA	Intervention DAX Copilot a été utilisé pour produire une note complète ou une section partielle de note (ex : histoire de la maladie actuelle, examen physique, évaluation et plan).	N = 46 Âge : non rapporté Proportions hommes/ femmes: non rapportées
		Comparateur/ Condition (s) de comparaison N/A	Spécialités : 17 spécialités médicales différentes Années d'expérience (moyenne) : 11,1 ans de pratique
Galloway 2024 États-Unis	Devis Étude pré-post sans groupe témoin	Intervention Documentation ambiante basée sur l'IA générative (outil non spécifié).	N = 31 Âge : 48,4 % âgés de 20 à 39 ans
	Temps de mesure T0: avant; T1: 60 jours après l'utilisation de l'IA	Comparateur/ Condition (s) de comparaison N/A	58 % hommes Spécialités : cliniciens en soins primaires Années d'expérience (moyenne): 8,4 ans
Harberle 2024 États-Unis	Devis Étude pré-post avec groupe témoin	Intervention DAX Copilot a été utilisé pour générer automatiquement des notes cliniques à partir des consultations. Un modèle standard de note a été utilisé.	GI: N= 44; GT: N= 43 Âge: non rapporté

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
	Temps de mesure T0: avant; T1: 8 mois après l'utilisation de l'IA	Une première révision et édition des notes a été effectuée par l'équipe DAX, puis une seconde révision par le clinicien, au besoin. Comparateur/ Condition (s) de comparaison Pratique habituelle de documentation (sans IA)	Proportions hommes/ femmes: non rapportées Spécialités: soins primaires adultes et pédiatriques (56,6 %), orthopédie et médecine du sport (15,2 %), spécialités chirurgicales, cardiaques, interventionnelles, neurosciences, obstétrique/gynécologie, oncologie et santé des femmes (11,1 %), spécialités médicales telles que l'endocrinologie et la rhumatologie (5,1 %) et psychiatrie (1 %). Années d'expérience: non rapportées
Kakaday 2025 États-Unis	Devis ECR Temps de mesure T0: avant ; T1: 3 mois après l'utilisation de l'IA	Intervention DAX Copilot a été utilisé pour générer des notes cliniques à partir des consultations. Un modèle standard de note a été utilisé. Comparateur/ Condition (s) de comparaison Pratique habituelle de documentation (sans IA)	GI N= 25 Âge: 44 % entre 36 et 55 ans 24 % hommes Spécialités: médecins (68 %) et infirmières praticiennes (32 %) en médecine de famille, médecine interne, pédiatrie, soins d'urgence
			Années d'expérience (moyenne): 8,7 ans* (significativement différent du GT) GT N= 20

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
Liu 2024 États-Unis	Devis Étude pré-post avec groupe témoin Temps de mesure T0: avant ; T1: 5 semaines après l'utilisation de l'IA	Intervention DAX Copilot Aucune information supplémentaire n'a été fournie. Comparateur/ Condition (s) de comparaison Pratique habituelle de documentation (sans IA)	Âge: 70 % entre 36 et 55 ans 30 % hommes Spécialités: médecins (55%) et infirmières praticiennes (45%) en médecine de famille, médecine interne, pédiatrie, soins d'urgence Années d'expérience (moyenne): 7,5 ans GI N = 85 Âge: non rapporté 51 % hommes Spécialités: cliniciens en médecine familiale, médecine interne et pédiatrie générale (médecins et praticiens avancés) Années d'expérience: 39 % entre 15 et 25 ans; 38 % entre 5 et 15 ans GT N = 55 Âge: non rapporté 31 % hommes

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
			Spécialités : lieux de pratique et spécialités similaires à celles du groupe intervention Années d'expérience : 35 % entre 15 et 25 ans ; 42 % entre 5 et 15 ans
Ma Stephen 2025 États-Unis	Devis Étude pré-post sans groupe témoin	Intervention Dax Copilot a été utilisé pour générer une note complète à partir d'un modèle	N= 45 Âge: non rapporté
	Temps de mesure T0: avant ; T1: 3 mois après l'utilisation de l'IA	standardisé structuré en quatre sections : histoire de la maladie actuelle, examen physique, évaluation et plan.	38 % hommes Spécialités: soins primaires, rhumatologie,
		Comparateur/ Condition (s) de comparaison N/A	cardiologie, gastrologie, gériatrie, ophtalmologie, orthopédie, otorhinolaryngologie
			Années d'expérience: 38 % > 15 ans; 31 % entre 10 et 14 ans
Misurac 2025	Devis Étude pré-post sans groupe	Intervention Un outil d'IA ambiante permettant de	N=35
États-Unis	témoin	transcrire automatiquement les conversations patient-clinicien et de	Âge: 57 % entre 35 et 44 ans
	Temps de mesure	produire des notes cliniques préliminaires. Ces notes sont ensuite	23 % hommes
	T0: avant ; T1: 5 semaines après l'utilisation de l'IA	revues par les participants, corrigées au besoin, puis intégrées dans le DME.	Spécialités: médecine de famille, médecine interne, médecine pédiatrique, gynécologie-obstétrique, oncologie, médecine de sport et soin d'urgence
		Comparateur/ Condition (s) de comparaison N/A	Années d'expérience: 48 % > 10 ans

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
Nguyen 2023 États-Unis	Devis Étude mixte - pré-post sans groupe témoin Temps de mesure T0: avant ; T1: 1 mois après l'utilisation de l'IA	Intervention DAX Copilot (versions 3.0.5, 3.0.6, et 3.0.7) a été utilisé pour générer automatiquement des notes cliniques à partir des consultations. Un modèle standard de note a été utilisé. Une première révision et édition des notes a été effectuée par l'équipe DAX, puis une seconde révision par le clinicien, au besoin.	Age (moyenne): 46,6 ans 43 % hommes Spécialités: médecins et infirmières en endocrinologie (33,3 %), oncologie médicale gastro-intestinale (11,1 %), conseil génétique (11,1 %), tête et cou (11,1 %), sarcome
		Comparateur/ Condition (s) de comparaison N/A	(22,2 %) et médecine interne (11,1 %) Années d'expérience: non rapportées
Omon 2025	Devis Série temporelle interrompue de type AB: A, phase contrôle et B,	Intervention Medimo (Pleap, Inc.), une application web qui permet de transcrire les entrées	N = 12 Âge (moyenne): de 33,2 ans
Japon	phase d'intervention Temps de mesure Chaque phase a duré 5 jours	vocales des professionnels de santé pendant les consultations, puis de résumer le contenu. Les cliniciens peuvent consulter et éditer le résumé avant de l'enregistrer dans le DME.	33% hommes Spécialités: physiothérapeutes, ergothérapeutes, orthophonistes
		Comparateur/ Condition (s) de comparaison Pratique habituelle de documentation (sans IA)	Années d'expérience: 10,1 ans.
Owens 2024	Devis	Intervention Dax Copilot	N = 83 ont répondu à l'enquête (75%); 19 ont été inclus dans les analyses avant-après

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
États-Unis	Enquête transversale + une étude pré-post rétrospective sans groupe témoin	Aucune information supplémentaire n'a été fournie.	Âge (moyenne) : 43,5 ans
		Comparateur/ Condition (s) de	35 % hommes
	Temps de mesure	comparaison	
	Temps de mesure non précisé	N/A	Spécialités : Médecine interne et médecine de famille
			Années d'expérience: 13,8 ans
Shah,	Devis	Intervention	N=22
Crowell, et	Étude qualitative	Dax Copilot a été utilisé pour générer	
al.,		une note complète à partir d'un modèle	Âge: non rapporté
2025	Temps de mesure	standardisé structuré.	
États-Unis	Un seul temps de mesure, après		Proportions hommes/ femmes: non
	3 mois d'implantation de DAX	Comparateur/ Condition (s) de	rapportées
		comparaison	
		N/A	Spécialités: 13 médecins en soins primaires et 9 en spécialités ambulatoires
			9 en specialites ambulatoires
			Années d'expérience: non rapportées
Shah, Devon-	Devis	Intervention	N = 38
Sand, et al.,	Étude pré-post sans groupe	Dax Copilot a été utilisé pour générer	
2025	témoin	une note complète à partir d'un modèle	Âge: 40 % entre 45 et 54 ans
États-Unis		standardisé structuré en quatre	
	Temps de mesure	sections : histoire de la maladie actuelle,	60 % hommes
	T0: avant et T1: après 3 mois d'utilisation	examen physique, évaluation et plan.	Spácialitás : Soins primairas (CC 0/) at
	a utilisation	Comparatour / Condition (c) do	Spécialités : Soins primaires (66 %) et
		Comparateur/ Condition (s) de comparaison	spécialités ambulatoires (34 %)
		N/A	Années d'expérience : 66 % ont plus de 10
			ans

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
Stults 2025 États-Unis	Devis Étude pré-post sans groupe témoin Temps de mesure T0: avant et T1: après 3 mois d'utilisation	Intervention Une plateforme d'IA ambiante a été utilisée pour générer automatiquement des notes de progression à partir des consultations. Comparateur/ Condition (s) de comparaison N/A	N = 57 64 % hommes Âge (moyenne): 50,3 ans Spécialités: soins primaires (66,7 %), surspécialités médicales (19,3 %) et surspécialités chirurgicales (14 %) Années d'expérience: 12,7 ans ± 8,0
Van Buchem 2024 Pays-Bas	Devis Étude intra-sujet Chaque étudiant a d'abord résumé 4 simulations de consultations manuellement, puis 8 avec Autoscriber (incluant l'enregistrement et la modification du résumé), et enfin 4 consultations manuellement.	Intervention Autoscriber, une application web qui transcrit et résume les conversations médicales en utilisant un mélange de GPT-3.5 et GPT-4. Comparateur/ Condition (s) de comparaison Génération de résumé manuelle	N=21 étudiants ayant un doctorat en médecine Les caractéristiques de l'échantillon ne sont pas rapportées
Wang 2021 États-Unis	Temps de mesure N/A Devis Étude intra-sujet Chaque participant a révisé la note générée par l'outil de	Intervention Un système de documentation automatique centré sur le patient utilisant la reconnaissance vocale et le	N=2 étudiants en médecine. Chaque étudiant a réalisé 32 consultations simulées Âge: 24 et 34 ans.

1 ^{er} Auteur Année Pays	Devis Temps de mesure	Intervention Comparateur/ Condition (s) de comparaison	Échantillon
	documentation numérique, puis a resaisi et dicté ce même contenu, permettant ainsi une comparaison intra-sujets entre les trois méthodes	traitement automatique de langage naturel. Le système permet de résumer le contenu des simulations de consultations et placer les informations dans la section correspondante de la note médicale en utilisant les indices fournis par les cliniciens.	Les autres caractéristiques de l'échantillon ne sont pas rapportées.
	Temps de mesure N/A	Comparateur/ Condition (s) de comparaison	
		Saisie manuelle Les participants ont tapé manuellement les notes cliniques générées par l'outil d'IA.	
		Dictée Les participants ont dicté les notes cliniques générées pas l'outil d'IA en utilisant un outil de reconnaissance vocale (Google Chrome Web Speech	
		API). Les notes dictées étaient transcrites automatiquement, puis révisées et corrigées par les participants.	

ECR : essai clinique randomisé; GI : groupe d'intervention; GT : groupe témoin; IA : intelligence artificielle; N/A : non applicable, N.B. : notez bien

ANNEXE 5. VARIABLES D'INTÉRÊT ET OUTILS DE MESURE

1 ^{er} Auteur	Variables et outils de mesure
Albercht 2025	 La productivité perçue : évaluée dans les sondages pré et post-intervention La charge de documentation clinique: évaluée seulement dans les sondages post-intervention Le travail effectué en dehors des heures régulières: évalué seulement dans les sondages post-intervention L'épuisement professionnel des cliniciens: évalué seulement dans les sondages post-intervention La satisfaction au travail: évaluée seulement dans les sondages post-intervention Toutes les variables ont été mesurées avec un sondage élaboré par deux médecins spécialisés en informatique médicale et un analyste d'affaires. Les réponses aux questions du sondage ont été enregistrées sur des échelles de Likert de 5 points, allant de « Tout à fait en désaccord » à « Tout à fait d'accord ».
Baker 2024	 Temps de documentation: temps écoulé à partir du premier point d'interaction avec la méthode d'enregistrement Efficience de la documentation: qualité/temps La qualité de la documentation a été mesurée avec le Documentation Quality Instrument (PDQI-9) qui évalue la précision, l'exhaustivité et l'organisation de la documentation des médecins.
Balloch 2024	 L'expérience des cliniciens en matière de documentation, avec et sans l'outil d'IA, a été évaluée à l'aide d'un questionnaire personnalisé de huit items utilisant une échelle de Likert en cinq points. La charge de travail perçue a été mesurée pour les deux outils à l'aide du NASA Task Load Index (six items, échelle de 1 à 10). Le temps total de consultation : calculé comme étant le temps écoulé entre l'entrée et la sortie du patient (minutes) Le temps de la conversation clinicien-patient: calculé comme étant le temps consacré aux échanges entre le clinicien et le patient
	Pour chaque clinicien, les moyennes ont été calculées à partir de trois consultations réalisées avec et sans l'outil d'IA.
Barak- Corrren 2024	 Temps de documentation (avec et sans ChatGPT): Temps depuis le premier caractère jusqu'au dernier, ou depuis le début de la consultation du résumé généré par ChatGPT jusqu'au dernier caractère Effort perçu (échelle de Likert de 0 à 10) Efficience des résumés générés par ChatGPT : échelle de Likert de 0 à 10

Bundy	Expérience des cliniciens: Entrevues semi-dirigées
2024	
Duggan	 Temps consacré aux notes par rendez-vous
2025	Le nombre de dossiers fermés le jour même
	 Le temps de travail en dehors des heures normales
	La charge perçue liée à la documentation clinique
	Toutes les variables listées ci-dessus ont été mesurées à l'aide d'un outil d'analytique du dossier médical électronique.
	 Expérience des cliniciens: sondage avec des réponses enregistrées sur une échelle de Likert (0 à 7) et une question ouverte
Galloway 2024	Bien-être des cliniciens : comparaison de la proportion de réponses négatives au sondage
	Le sondage du post-test comprenait des questions supplémentaires portant sur la probabilité de recommander la technologie, son impact sur la productivité ainsi que les intentions quant à son utilisation future.
Harberle	 Expérience des cliniciens (engagement, sécurité, résilience et décompression et équilibre travail-vie personnelle) :
2024	mesurée à l'aide du sondage Press Ganey Workforce Engagement (6 items notés sur une échelle de Likert à 5 points).
	• Productivité : suivie à partir des work relative value units (wRVU), un indicateur standardisé qui reflète la quantité et la
	complexité du travail clinique réalisé. Des rapports opérationnels standards (mensuels, cumulés et annualisés par rapport
	aux cibles) ont été utilisés.
	Panel VBC : mesuré par le nombre de patients attribués à chaque médecin dans le cadre du système de soins basés sur
	la valeur (<i>Value-Based Care, VBC</i>), via les rapports de Castell.
	Temps de documentation par patient: extrait des rapports du DME.
	 Temps en dehors des heures de travail (activité entre 18h et 6h): extrait des rapports du DME.
	 Délais de documentation dans les 24h: proportion de consultations documentées dans les 24 h suivant la visite, extraite des rapports mensuels.
	 Les délais de soumission de codes de facturation: proportion de consultations facturées dans les 24 h suivant la visite,
	extraite des rapports mensuels.
Kakaday	Temps de documentation: défini comme le nombre de minutes consacrées à la rédaction des notes par visite. Mesuré
2025	à partir des journaux d'activité des cliniciens.
Liu	Expérience des cliniciens: sondage comportant des questions sur le temps passé sur le DME à domicile, le temps
2024	hebdomadaire consacré au DME en dehors des heures normales de travail, le temps de documentation après les
	consultations et la frustration liée à l'utilisation du DME. Les réponses ont été enregistrées sur une échelle de 1 à 5.

Ma Stephen	Temps par note: calculé comme une moyenne pondérée par période pour donner plus de poids aux périodes où le
2025	 Temps par note: calculé comme une moyenne pondérée par période pour donner plus de poids aux périodes où le clinicien a écrit plus de notes
2025	Temps quotidien consacré à la documentation
	 Temps quotidien consacre à la documentation Temps quotidien passé sur le DME en dehors des heures habituelles
	 Temps quotidien passé sur le DME Temps total quotidien passé sur le DME
	Toutes les variables sont extraites de l'application liant Copilot au DME (EPIC)
Misurac 2025	• L'épanouissement et l'épuisement professionnel ont été mesurés à l'aide du Stanford Professional Fulfillment Index (PFI), un outil qui évalue trois dimensions : l'épanouissement professionnel, l'épuisement lié au travail et le
	désengagement interpersonnel. Le score global d'épuisement est calculé en additionnant les sous-scores d'épuisement au travail et de désengagement interpersonnel, puis en les normalisant sur une échelle de 0 à 10, où des valeurs plus faibles indiquent un niveau accru d'épuisement.
	 Satisfaction des cliniciens: évaluée avec un sondage comprenant plusieurs questions ouvertes portant sur les interactions avec les patients, l'équilibre travail-vie personnelle et la satisfaction générale vis-à-vis de l'outil d'IA ambiante.
Nguyen	Volet quantitatif
2023	• Bien être des cliniciens: des facteurs de stress liés au travail incluant le niveau d'épuisement, le temps perçu comme suffisant pour la documentation et l'autonomie professionnelle ont été évalués à l'aide du Mini-Z (10 items) envoyé sous forme de sondage.
	 La qualité du sommeil a été mesurée au moyen du questionnaire Patient-Reported Outcomes Measurement Information System Short Form 4a Sleep Disturbance measures.
	Volet qualitatif
	• Expériences des cliniciens: entrevues semi-dirigées comportant des questions sur l'utilisation de l'outil d'IA en clinique et son impact sur le flux et la charge de travail.
Omon 2025	• Temps de documentation: mesuré par un tiers, à partir du moment où le participant était assis devant un ordinateur avec le DME prêt pour la documentation, jusqu'à l'enregistrement final de la note clinique.
	Le temps de documentation a été mesuré quotidiennement, et la moyenne pour chaque phase a ensuite été calculée. • Charge de travail: mesurée à l'aide du NASA Task Load Index (NASA-TLX).
Owens 2024	• Épuisement: mesuré avec le <i>Oldenburg Burnout Inventory</i> (OLBI) qui comprend deux sous-scores : le désengagement (OLBI-D) et l'épuisement (OLBI-E), chacun constitué de 8 items évalués sur une échelle de Likert (1 à 4)

	 Temps de documentation: correspond à la différence du temps de documentation (en minutes) par note et par consultation. Pour l'analyse de cette variable, seuls les participants à fort usage de DAX et disposant d'au moins 3 mois de données avant et après l'implémentation de DAX™ ont été inclus.
	Les variables secondaires comprenaient : • Temps consacré à la documentation en dehors des heures de travail habituelles (minutes par jour)
Shah, Crowell, et al., 2025	• Expérience des cliniciens: les dimensions évaluées incluent l'exigence cognitive et temporelle, la charge de travail et l'intégration travail-vie personnelle, les changements dans les flux de travail et l'engagement du médecin avec les patients. L'évaluation a été réalisée au moyen d'entretiens semi-structurés.
Shah, Devon- Sand, et al., 2025	 Charge liée au DME : évaluée à l'aide le <i>Physician Task Load</i> (PTL, 4 items) dérivé d'une version adaptée du NASA-TLX. Épuisement : mesuré avec la sous-échelle « épuisement au travail » du <i>Stanford Professional Fulfillment Index</i> (PFI) Gain de temps perçu : mesuré dans l'enquête post-intervention à l'aide d'une échelle allant de 90 minutes gagnées à 90 minutes perdues par demi-journée clinique.
Stults 2025	 Épuisement : évalué par la question spécifique sur l'épuisement du Mini-Z de l'American Medical Association. Charge de travail : mesurée par les dimensions « demande mentale », « demande temporelle » et « effort » du NASA-TLX. Temps hebdomadaire consacré à la documentation en dehors des heures de clinique. Capacité à accorder une attention exclusive aux patients pendant les consultations.
Van Buchem 2024	Temps de rédaction (secondes)
Wang 2021	Vitesse de rédaction des notes (mots/minute)

DME: dossier médical électronique; IA: intelligence artificielle; NASA-TLX: NASA Task Load Index; OLBI: Oldenburg Burnout Inventory; PDQI: Documentation Quality Instrument; wRVU: Work Relative Value Units; PDQI-9: Documentation Quality Instrument; PFI: Professional Fulfillment Index; PTL: Physician Task Load; VC: Value-Based Care

ANNEXE 6. ANALYSES STATISTIQUES, LIMITES MÉTHODOLOGIQUES ET RÉSULTATS

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
Albercht 2025	Analyses descriptives Rapports de côte Analyses de covariables	 Stratégies d'échantillonnage non spécifiées / échantillon non représentatif Les questions du sondage ne sont pas validées Les questions du sondage préimplantation ne sont pas exactement les mêmes que celles du sondage postimplantation Les participants au sondage préimplantation ne sont pas exactement les mêmes que ceux du sondage postimplantation et en raison de l'anonymat du sondage préimplantation, il n'a pas été possible d'apparier les réponses individuelles entre les périodes pré et post 	Flux de travail: amélioration significative après l'implantation de la plateforme d'IA ambiante. Les cliniciens étaient près de 7 fois plus susceptibles de juger la documentation facile (IC: 3,90–12,56; p < 0,001) et 5 fois plus susceptibles d'indiquer pouvoir compléter leurs notes avant la visite suivante (IC: 2,87–8,69; p < 0,001). La charge de documentation clinique: la majorité des répondants ont indiqué que la plateforme avait rendu la documentation facile (81%), permis de compléter la note avant la visite suivante (43%) et diminué la charge de documentation (77%) Le travail effectué en dehors des heures régulières: 73% des répondants ont indiqué que la plateforme avait réduit le temps de documentation en dehors des heures cliniques, Épuisement professionnel: 99,7% ont indiqué que la plateforme a diminué le risque d'épuisement professionnel lié à la documentation Satisfaction des cliniciens: 64% ont indiqué que la plateforme a augmenté la satisfaction. De plus, 47% ont estimé qu'une rencontre avec un patient supplémentaire pourrait être ajoutée si nécessaire.

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			Les effets observés étaient similaires, quel que soit le type de spécialité ou la durée d'utilisation de la plateforme.
Baker 2024	ANOVA pour mesures répétées	 Échantillon non représentatif (N=11) Facteurs de confusion non pris en considération La validité de l'outil utilisé pour évaluer la qualité de la documentation est compromise L'utilisation d'histoires de patients standardisées peut ne pas refléter la complexité et la variabilité des véritables antécédents des patients, ce qui limite la validité externe de l'étude. 	Temps de documentation : la dictée était la méthode de documentation la plus rapide, significativement plus rapide que la saisie au clavier (temps moyen de 43 secondes contre 96 secondes, respectivement ; P = 0,001 ; IC 95 % : 20,9 à 85,3 ; Figure 1). ChatGPT était intermédiaire en termes de rapidité, avec un temps moyen de 69 secondes, mais cette différence n'était pas significative par rapport à la dictée (P = 0,14 ; IC 95 % : -6,1 à 258,3) ou à la saisie au clavier (P = 0,12 ; IC 95 % : -25,2 à 59,2). Efficience (PDQI-9/min) : Le score d'efficience de la dictée était significativement supérieur à celui de la saisie au clavier (P < 0,001), tandis qu'aucune différence significative n'a été observée entre la dictée et ChatGPT (P = 0,06).
Balloch 2024	Analyses descriptives Tests non paramétriques	 Échantillon non représentatif (n=7) Facteurs de confusion non pris en considération L'utilisation de scénarios patients fictifs pourrait ne pas refléter des consultations très complexes avec de larges volumes d'informations médicales préexistantes. 	Expérience des cliniciens: Une amélioration de l'expérience des cliniciens a été rapportée. 100 % des cliniciens ont indiqué pouvoir accorder toute leur attention aux patients lorsqu'ils utilisaient l'outil d'IA, contre 66 % sans l'outil, et 94 % ont rapporté une expérience positive en clinique, contre 71 % sans l'outil d'IA. Une réduction de la charge de travail perçue par le clinicien a également été observée sur 5/6 dimensions du NASA-TLX lors de l'utilisation de l'outil d'IA.

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			Le temps total de consultation : L'utilisation de l'outil d'IA a permis de réduire significativement la durée totale des consultations (-193 s, soit 26,3 % de gain de temps ; p = 0,03).
			Le temps de la conversation clinicien-patient: Le temps moyen consacré aux échanges avec les patients et leur famille n'a pas changé significativement (28 s ; p = 0,32).
Barak- Corrren 2024	Moyenne et intervalles de confiance	 Échantillon non représentatif (N=10, tous issus de la même organisation médicale et travaillant en médecine d'urgence pédiatrique) Seulement des analyses descriptives 	Temps de documentation: La rédaction des notes de supervision médicale a été 36 à 43 % plus rapide avec ChatGPT pour deux cas cliniques, tandis qu'aucune différence notable n'a été observée pour les deux autres.
		 Portée limitée des scénarios cliniques inclus dans l'étude 	Effort perçu : était plus faible pour un seul cas (fièvre d'origine inconnue), passant de 6,6/10 à 4,4/10 (-33%), tandis que les autres cas ne présentaient pas de différences.
			Efficience des résumés générés par ChatGPT: était notée 8,2/10
Bundy 2024	Analyses inductives	 Devis qualitatif La relation entre les chercheurs et les participants n'a pas été explicitement abordée dans l'étude, ce qui ne permet pas d'évaluer dans quelle mesure cette relation a pu influencer la collecte ou l'interprétation des données. 	Expérience des cliniciens: La majorité des médecins interrogés ont rapporté que DAX réduisait le temps consacré à la documentation clinique, allégeait la charge cognitive et facilitait des consultations plus engageantes. Plus précisément, les participants ont mentionné une amélioration de leur qualité de vie (n = 10), une réduction significative du travail effectué en dehors des heures régulières (n = 7), un allègement de la charge cognitive liée aux tâches cliniques (n = 6), ainsi qu'une plus grande attention et

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			une interaction plus personnalisée lors des rencontres avec les patients (n = 4).
Duggan 2025	Une modélisation à effets aléatoires a été utilisée pour contrôler les covariables comme le rôle du clinicien, la spécialité médicale, la complexité des patients, la charge de travail du clinicien et les années d'expérience du clinicien.	 Échantillon non représentatif (stratégie d'échantillonnage biaisée) Courte durée d'intervention (selon les auteurs) Les résultats s'appuient sur les perceptions des cliniciens concernant leurs interactions avec les patients, plutôt que sur une mesure directe de l'expérience des patients 	Temps de documentation: L'utilisation de Dax a été associée à une réduction de 20,4 % du temps consacré aux notes par rendez-vous par rapport à la situation initiale (P < 0,001). Le nombre de dossiers fermés par jour (taux de clôture des rendez-vous le jour même) était supérieur de 9,3 % P < 0,001); Le temps hebdomadaire de travail en dehors des heures normales par jour de travail était réduit de 30,0 % (P = 0,02);
	Test t Analyse thématique (pour la question ouverte)		Les résultats sont restés significatifs après contrôle pour la spécialité du clinicien, la complexité des patients, la charge de travail du clinicien et l'expérience du clinicien.
			Expérience des cliniciens: Les résultats du sondage ont montré que l'utilisation de DAX a permis significativement de réduire la distraction des cliniciens liée à la documentation, de diminuer leur surcharge mentale et le sentiment d'épuisement associé à la charge de documentation. Il a également permis aux cliniciens d'être plus efficaces dans la rédaction des notes, de consigner autant de détails qu'ils le souhaitaient et de suivre le volume actuel de leurs patients. Dans leurs réponses ouvertes, les cliniciens ont mentionné une plus grande efficacité dans leur flux de travail ainsi qu'une réduction du temps consacré à

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
Galloway 2024	Test t pour échantillons appariés	 Échantillon non représentatif Facteurs de confusion non pris en considération Les choix de réponse aux questions ont été modifiés entre les enquêtes pré- et post-intervention, ce qui a pu influencer la manière dont certaines questions ont été interprétées 	la rédaction des notes et une diminution de la durée des consultations. Ils ont également indiqué que les notes générées par la transcription ambiante n'éliminaient pas complètement la charge de la documentation clinique. Toutefois, ils ont souligné que l'utilisation de DAX réduisait l'effort mental requis, en enregistrant les détails difficiles à mémoriser et en supprimant la tâche difficile de rédiger une note entièrement à partir de zéro. Les cliniciens ont indiqué que l'utilisation de DAX leur permettait de mieux interagir avec les patients, sans être distraits par la saisie au clavier Bien-être des cliniciens: 71 % des participants ont rapporté un impact négatif lié à leur processus de documentation habituel, comparativement à 38,7 % après l'intervention (p = 0,010).
Haberle 2024	Analyses de régression (linéaire, logistique et de Poisson) Test t pour échantillons appariés	Aucune information sur la validité et la fidélité des outils de mesure	Expérience des cliniciens: Les participants du GI ont obtenu des scores plus élevés pour l'engagement, la sécurité, la résilience/décompression et l'équilibre travail-vie personnelle (on ne précise pas si les différences sont significatives) Productivité: La productivité (wRVN) a augmenté de 1% significativement dans le GI (p < .001), sans aucun changement pour GT. Le nombre de patients attribués à chaque médecin dans le cadre du système de soins basés sur la

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			 valeur: aucun changement significatif pour le GI ou le GC. Utilisation du DME: Le GI passait en moyenne 14,2 % de son temps en dehors des heures régulières dans le DME, comparé à 14,9 % dans le GT. En fin d'étude, ce temps a significativement augmenté dans le GI (+4,69 %, p < .05) mais diminué dans le GT (-0,95 %). Le temps de documentation par patient a significativement diminué dans le GI, sans aucun changement pour le GT. Les délais de documentation dans les 24 h ont baissé significativement dans les deux groupes sans différence significative entre le groupe d'intervention et le groupe témoin. Les délais de soumission de codes de facturation ont augmenté significativement dans les deux groupes (p< 0,001) mais l'utilisation de DAX n'a pas entraîné de différence significative par rapport au contrôle.
Kakaday 2025	Test de Kruskal Wallis Test Wilcoxon	 Échantillon non représentatif La création de sous-groupes a réduit la taille d'échantillon et a empêché les comparaisons inter- groupes Les participants et les évaluateurs n'étaient pas à l'aveugle. 	*Des sous-groupes ont été constitués en fonction de la fréquence d'utilisation de DAX et l'analyse a porté sur les différences au sein de ces sous-groupes dans le temps, plutôt que sur les différences entre les groupes assignés aléatoirement. : Utilisateurs fréquents (< 45 % de toutes les visites; N = 7) Utilisateurs modérés (45 – 69,9 % de toutes les visites; N = 6)

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			Utilisateurs faibles (> 70 % de toutes les visites; N = 12) Temps de documentation : Aucun changement significatif pour aucun des sous-groupes
Liu 2024	Tests Khi-carré	 Biais de sélection potentiels Facteurs de confusion non pris en considération Biais de rappel potentiels Questions du sondage n'ont pas été validées 	Expérience des cliniciens: Dans le groupe intervention, 47,1 % ont déclaré une diminution du temps passé sur le DME à domicile, contre 14,5 % dans le groupe contrôle (<i>P</i> < 0,001). De même, 44,7 % ont rapporté une diminution du temps hebdomadaire consacré au DME en dehors des heures normales de travail, contre 20,0 % dans le groupe contrôle (<i>P</i> = 0,003). En outre, 43,5 % ont indiqué une réduction du temps de documentation après les consultations, comparativement à 18,2 % dans le groupe contrôle (<i>P</i> = 0,002). De plus, 44,7 % ont signalé une diminution de la frustration liée à l'utilisation du DME, contre 14,5 % dans le groupe contrôle (<i>P</i> < 0,001). À l'inverse, en moyenne, 44,7 % des participants du groupe intervention et 68,7 % du groupe contrôle ont jugé que leur expérience du DME était comparable avant et après l'intervention.
Ma Stephen 2025	Test t	 Échantillon non représentatif Aucune information sur la validité des mesures Facteurs de confusion non pris en considération 	Le temps médian par note a diminué significativement et est passé de 4,86 minutes (IQR 2,89–6,47) en période pré-intervention à 3,64 minutes (IQR 2,45–5,46) en période d'intervention (P < 0,001). Le temps quotidien consacré à la documentation a diminué de 6,89 minutes (IQR –22,37 à –0,65), également de façon significative (P < 0,001).

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			Temps quotidien passé sur le DME en dehors des heures habituelles : une réduction significative de 7,17 minutes (P = 0,01) Temps quotidien passé sur le DME: une réduction significative de 19,95 minutes (P < 0,001).
Misurac 2025	Test des signes pour échantillons appariés	 Échantillon non représentatif Facteurs de confusion non pris en considération 	Épuisement professionnel: Une réduction significative ($p = 0.005$). Le taux d'épuisement a diminué de 69 % (24/35) à 43 % (15/35). En subdivisant l'échantillon en quartiles selon le niveau d'utilisation de l'IA, on montre que le quartile le plus bas d'utilisation ne présentait aucune amélioration du score d'épuisement contrairement aux trois quartiles supérieurs qui présentaient une réduction des scores d'épuisement. Dimension sur l'épuisement lié au travail: Aucune amélioration significative n'a été observée. Dimension sur le désengagement interpersonnel: une amélioration significative ($p < 0.001$). Dimension sur l'épanouissement professionnel: aucune différence significative n'a été constatée ($p = 0.10$).
			Satisfaction des cliniciens: 28/35 participants ont rapporté un impact positif sur les interactions avec les patients, attribuable à l'outil d'IA. Ils ont notamment souligné une amélioration de l'écoute et de l'attention lors des consultations. De plus, 24 /35 ont indiqué une amélioration de leur équilibre travail-vie personnelle, expliquant ce bénéfice non seulement par un le gain de temps, mais

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			surtout par une réduction du stress lié à la documentation
Nguyen 2023	Test t pour échantillons appariés Analyse thématique déductive et inductive	 Échantillon non représentatif Manque de puissance statistique pour détecter des changements significatifs dans le bien âtre des cliniciens Biais d'attrition potentiels Facteurs de confusion non pris en considération 	Bien être des cliniciens Aucune différence significative n'a été observée dans les scores composites du Mini-Z ($p = 0,257$). Il n'y avait pas de différence significative dans l'épuisement ($p = 0,081$) ni dans la qualité du sommeil ($p = 0,195$). Temps perçu comme suffisant pour la documentation: amélioration significative (2,1 vs 3,6; $p = 0,005$). Charge de travail Tous les cliniciens ont indiqué avoir dû éditer les notes générées avant de les signer, avec une intensité allant de modifications mineures à majeures. Toutefois, la charge d'édition s'est allégée au fil du temps.
Omon 2025	Tests paramétriques et non paramétriques	 Échantillon non représentatif (N=12) Durée d'intervention/observation très courte (5 jours) Facteurs de confusion non pris en considération 	Temps de documentation: Le temps moyen de documentation était de 305 ± 53,9 secondes durant la phase A (documentation habituelle) et de 336 ± 78,0 secondes durant la période B (utilisation de l'IA), sans différence significative entre les deux périodes. Charge de travail: aucune différence significative en ce qui concerne la demande mentale, la demande physique, la performance perçue ou l'effort. En revanche, la pression temporelle était significativement plus faible lors de l'utilisation de l'IA

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			(p = 0,037), tandis que le niveau d'insatisfaction était significativement plus élevé (p = 0,041).
Owens 2024	Tests t Régressions linéaires Modèles à effets mixtes	 Échantillon non représentatif et limites de généralisabilité des résultats à d'autres cohortes de soins primaires Certains facteurs de confusion n'étaient pas identifiés et pris en considération 	Épuisement (OLBI): Une utilisation élevée de DAX était associée à un niveau significativement plus faible d'épuisement sur le sous-score de désengagement de l'OLBI (différence moyenne = −2,1 ; IC à 95 % : −3,8 à −0,4). En revanche, aucune différence significative n'a été observée pour le sous-score d'épuisement ni pour le score total. Temps moyen de documentation par note : une réduction significative de 28,8 % après l'implantation de DAX (−1,8 minute ; IC à 95 % : −2,2 à −1,4). Temps de documentation en dehors des heures planifiées : diminution significative de 11,8 % (−4 minutes ; IC à 95 % : −7,2 à −1,0).
Shah, Crowell, et al., 2025	Analyse thématique inductive et déductive	 Biais d'échantillonnage potentiels, puisque la majorité des médecins interrogés étaient des utilisateurs fréquents de l'outil d'IA Il manque d'information sur la relation entre le chercheur et les participants 	Expérience des cliniciens: Les médecins ont rapporté des perceptions positives concernant la demande cognitive (100 %), la demande temporelle (62 %), l'intégration travail-vie personnelle (91 %) et la charge de travail globale (89 %). Des commentaires positifs concernant l'engagement avec les patients ont été rapportés dans 68 % des cas.
Shah, Devon- Sand, et al., 2025	Tests t Tests de Wilcoxon Taille d'effet de Cohen	 Échantillon non représentatif (biais de sélection potentiels) Facteurs de confusion non pris en considération 	Charge liée au DME : réduction statistiquement significative du score total du <i>Physician Task Load</i> ($-24,42$ sur une échelle de 0 à 100 ; $p < 0,001$) Épuisement : réduction statistiquement significative ($-1,94$ sur une échelle de 0 à 10 ; $p < 0,001$).

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
			Gain de temps perçu (mesuré seulement en post intervention): Le gain de temps médian perçu par demi-journée de clinique était de 20 minutes
Stults 2025	Khi carré Test de Fisher Test t Mann Whitney	 Échantillon non représentatif Facteurs de confusion non pris en considération 	Épuisement: aucun changement significatif Charge de travail: Sur une échelle de 0 à 20, il y a une diminution significative pour la demande mentale (de 12,2 ± 4,0 à 6,3 ± 3,7), la demande temporelle (de 13,2 ± 4,0 à 6,4 ± 4,2) et l'effort requis pour rédiger les notes (de 12,5 ± 4,1 à 7,4 ± 4,3) (p < 0,001). Temps hebdomadaire consacré à la documentation: 54,4 % des répondants (n = 31) déclaraient passer une heure ou moins par semaine sur les notes en post-intervention, contre seulement 14,0 % (n = 8) avant (p < 0,001). Capacité à accorder une attention exclusive aux patients: la proportion de cliniciens ayant répondu « d'accord/très d'accord » est passée de 57,9 % (n = 33)
Van Buchem 2024	ANOVA Test de Tukey	 Échantillon non représentatif Facteurs de confusion non pris en considération L'utilisation des consultations simulées pourrait ne pas refléter fidèlement le caractère dynamique et souvent imprévisible des interactions cliniques en temps réel. 	à 93,0 % (n = 53) (p < 0,001). Temps de documentation : le temps médian nécessaire à la rédaction et à la modification du résumé avec Autoscribe était significativement inférieur à celui du résumé réalisé manuellement, soit 152 secondes [93–244] contre 202 secondes [128–286] (p = 0,02).

1 ^{er} Auteur	Analyses statistiques	Limites méthodologiques	Résultats
Wang 2021	ANOVA pour mesures répétées Test t pour échantillons appariés	 Échantillon non représentatif (N=2) Facteurs de confusion non pris en considération L'utilisation des consultations simulées pourrait ne pas refléter fidèlement le caractère dynamique et souvent imprévisible des interactions cliniques en temps réel. 	Temps de documentation: Pour les sections d'antécédents, l'outil de rédaction numérique était en moyenne 131,49 mots/minute plus rapide que la saisie (p < 0,0001) et 130,96 mots/minute plus rapide que la dictée (p < 0,0001). Aucune différence significative n'a été observée entre la saisie manuelle et la dictée (p = 1,00). Pour les sections examen physique, l'outil de rédaction numérique était en moyenne 59,41 mots/minute plus rapide que la saisie manuelle (p < 0,0001) et 74,94 mots/minute plus rapide que la dictée (p < 0,0001). La saisie manuelle était quant à elle 15,53 mots/minute plus rapide que la dictée (p < 0,0001).

DME : dossier médical électronique; GI : groupe d'intervention; GT : groupe témoin; IA : intelligence artificielle; IC : intervalle de confiance; IQR : écart interquartile; NASA-TLX : NASA Task Load Index; OLBI : Oldenburg Burnout Inventory; PDQI : Documentation Quality Instrument; wRVU : Work Relative Value Units