



L'Unité d'évaluation des technologies et des modes d'intervention en
santé (UETMIS) du Centre Universitaire de Santé McGill (CUSM)

Health Technology Assessment Unit (TAU) of the MUHC



February 6, 2025

Ambient AI Digital Scribes for Clinical Documentation

Rapid Environmental Scan and Vendor Comparison
Report no. 105

Report prepared for the Technology Assessment Unit (TAU) of the McGill University Health Centre (MUHC)

by

Nisha Almeida, PhD

**Manager and Scientific Lead, Health Technology Assessment Unit
(TAU), MUHC**

Mission Statement

The MUHC Health Technology Assessment Unit (TAU) advises hospital administrators and clinical teams in difficult resource allocation decisions. Using an approach based on independent, critical evaluations of the available scientific evidence and a transparent, fair decision-making process, novel and existing medical equipment, drugs and procedures used by healthcare professionals are prioritized on a continuous basis ensuring the best care for life with the best use of resources.

Declaration of Conflicts of Interest

Members of TAU's research staff and policy committee declare no conflicts of interest.

Suggested citation

Almeida N. Ambient AI Digital Scribes for Clinical Documentation: A Rapid Environmental Scan and Vendor Comparison. Montreal (Canada): Technology Assessment Unit (TAU) of the McGill University Health Centre (MUHC); 2025 February 6. Report no. 105. 28 pages

DISCLAIMER

The Technology Assessment Unit (TAU) of the McGill University Health Centre (MUHC) prepares evidence to inform institutional decision-making and, when appropriate, makes policy recommendations based on that evidence. This report draws on public literature, horizon-scanning reports, industry and institutional sources, and information available to the MUHC at the time of preparation.

The information should not be used as a substitute for professional clinical, legal, regulatory, procurement, or financial advice. Although care was taken to accurately summarize the source materials, technology maturity, regulatory status, and institutional use may change rapidly. The MUHC does not guarantee that the information is complete or current and is not responsible for errors, omissions, or losses arising from use or misuse of the information. Inclusion of a technology, company, or institution does not constitute endorsement.

EVIDENCE AT A GLANCE

WHAT DID WE EVALUATE?

Core functionalities, published evidence of clinical benefit and user experience, and privacy and implementation considerations of the following 9 ambient AI digital scribe solutions:

DAX Copilot · Nabla · Ambience · Scribeberry · NoteDx · Heidi · Coeurway · Tali AI · Plume AI

What does the evidence tell us?

OUTCOME	WHAT THE PUBLISHED EVIDENCE SHOWS	OVERALL SIGNAL
Documentation burden	Ambient scribes can reduce documentation burden, but gains vary - A large, controlled study of Nabla (303,266 encounters) found a statistically significant reduction in time spent writing notes both during the encounter and after hours. - DAX evaluations reported modest reductions in documentation time. - Time savings do not consistently translate into greater overall productivity or physician well-being. Benefits appear highly dependent on clinician workflow, editing needs and EHR integration.	SOME BENEFIT
Accuracy & note quality	Useful draft notes, but clinician review remains essential Ambient scribes can effectively summarize clinical encounters, particularly the history, but errors, omissions, incorrect attribution and hallucinated information have been reported.	MIXED
Physician well-being	Evidence for physician burnout and well-being is limited and inconsistent Mostly based on uncontrolled pre-post physician surveys.	MIXED
Patient experience	No clear improvement in patient experience has been demonstrated. - While some patients report more time spent talking with their doctors, others expressed unease about having visits recorded. - Potential benefit needs to be balanced against privacy/consent concerns.	EVIDENCE GAP

DECISION-LEVEL SIGNAL

Ambient AI scribes are a promising but still maturing technology. Evidence supports the potential to reduce documentation burden, but there is currently insufficient comparative evidence to identify a clearly superior vendor or to support organization-wide adoption without local evaluation.

TAU direction: conduct a limited, prospectively evaluated pilot measuring physician satisfaction, documentation burden, note accuracy, patient experience, privacy/security, EHR compatibility, and cost; maintain clinician review of all AI-generated documentation and develop continuous monitoring processes.

WHAT DIFFERENTIATES THE VENDORS

The basic ability to listen to an encounter and generate a clinical note is increasingly common across products. More meaningful differences are:

DIFFERENTIATOR	MARKET OBSERVATION	WHY IT MATTERS
1 EHR integration	- Direct integration may determine whether the technology actually improves workflow or simply introduces another application and copy-and-paste step. - DAX, Nabla and Ambience offered the most comprehensive integration, including EHR data retrieval and coding support	HIGH PRIORITY
2 Transcription quality & editing burden	- Peer-reviewed published evidence on note quality and accuracy exists specifically for DAX Copilot and Nabla. - The quality of the final note, not transcription accuracy alone, is critical. Performance during complex and lengthy encounters requires particular attention.	HIGH PRIORITY
3 Customization & Workflow	- Specialty-specific templates, note structure, level of detail and adaptation to individual physician documentation styles may influence adoption. - Scribeberry, a Canadian product, includes a talk-back function which allows the user to have a conversation with the scribe about the encounter.	TEST LOCALLY
4 Canada/Quebec fit	- Privacy & data architecture: All Canadian vendors were compliant with privacy/data residency laws. - Language: Recognition of local accents or expressions is also an advantage, which Quebec vendors (Plume AI and Coeurway) offer. - Data location, audio retention, third-party cloud and LLM providers, model training practices and compliance with Quebec requirements must be independently verified.	LOCAL ADVANTAGE
5 Evidence maturity	Published vendor-specific evidence was concentrated primarily on DAX Copilot and Nabla. For most other products, independent clinical outcome evidence was not identified in this scan.	HIGH PRIORITY

TABLE OF CONTENTS

Evidence at a glance.....	iii
What differentiates the vendors	iv
Table of Contents.....	v
List of Tables and figures	vi
List of Abbreviations	vii
Executive Summary.....	viii
Sommaire	xi
1. Background.....	1
2. Objective.....	1
3. Methods.....	1
3.1 Rapid environmental scan	1
3.2 Scope and limitations.....	2
4. Results.....	2
4.1 Vendor landscape and functionality	2
4.2 Published Evidence of Clinical and Operational Benefit.....	5
4.3 Canadian/Quebec Context.....	6
4.4 Cost and Value Considerations	7
5. Discussion: Implementation considerations	7
5.1 Evidence maturity	7
5.2 Note accuracy and quality	7
5.3 Patient experience	7
5.4 Canadian/Quebec Context.....	8
5.5 Cost and overall value.....	8
6. Conclusions.....	9
7. Recommendations.....	9
Appendices.....	11
References	14

LIST OF TABLES AND FIGURES

Table 1. Ambient AI scribe landscape*	3
Table 2. Main differences in vendor functionalities	4
Table 3. Evidence of clinical benefit	5

LIST OF ABBREVIATIONS

Abbreviation	Definition
AI	Artificial intelligence
EHR	Electronic health records
LLM	Large language models
MUHC	McGill University Health Centre
PIPEDA	Personal Information Protection and Electronic Documents Act
TAU	Technology Assessment Unit

EXECUTIVE SUMMARY

BACKGROUND

Ambient artificial intelligence (AI) digital scribes use speech recognition and generative AI to capture clinician–patient conversations and generate draft clinical documentation. These technologies have the potential to reduce documentation burden, improve clinical workflow, and allow clinicians to devote greater attention to patients during encounters. However, the market is evolving rapidly, with important differences between products in electronic health record (EHR) integration, functionality, privacy and data architecture, and limited independent evidence regarding their clinical and operational benefits.

OBJECTIVE

To compare ambient AI digital scribe solutions relevant to the Canadian and MUHC context by evaluating their functionalities, clinical benefits, strengths, limitations, and privacy considerations, with the aim of identifying feasible solutions that could improve documentation efficiency and clinician experience while meeting Quebec privacy and regulatory requirements.

METHODS

Environmental Scan

- We conducted a rapid environmental scan of nine ambient AI digital scribe solutions: DAX Copilot, Nabla, Ambience, Scribeberry, NoteDx, Heidi, Coeurway, Tali AI, and Plume AI.
- Products were compared according to core functionality, EHR integration and interoperability, language and customization capabilities, privacy and data-hosting considerations, cost and implementation requirements.
- The scan was supplemented by a targeted review of published evidence on clinical and operational outcomes and publicly available clinician user feedback. Vendor-reported claims regarding clinical outcomes were excluded from the assessment of effectiveness.

SCOPE AND LIMITATIONS

This was a rapid environmental scan rather than a systematic review or formal procurement evaluation. Vendor characteristics were derived from publicly available information, were not independently verified by TAU, and may change rapidly. Direct head-to-head comparative studies were not identified; consequently, findings should not be interpreted as establishing the clinical superiority of one product over another.

KEY FINDINGS

Core functionalities

- All nine products provided the core ability to capture clinical encounters and generate draft clinical documentation.
- The 3 largest vendors, **Dragon Ambient eXperience (DAX) Co-pilot** from Microsoft, **Nabla** (France) and **Ambience** (USA), were the most comprehensive in their support for diverse functionalities including integration with the electronic health record (EHR), ability to pull data from the EHR for viewing in the scribe app, and integrated ICD-10 coding assistance. Smaller and newer products more commonly relied on clinicians copying and pasting generated documentation into the EHR. Scribeberry, a Canadian product already being used at the MUHC, also offered a “talk-back” function allowing clinicians to query the scribe about the encounter.
- Canadian vendors may offer implementation advantages related to the local privacy environment. The scan identified five Canadian vendors (**Scribeberry, NoteDx, Tali AI, Coeurway, and Plume AI**) as reporting compliance with Canadian and Quebec privacy requirements.
- Quebec-developed products (**Coeurway, Plume AI**) may have the additional practical advantage of better recognizing local language expressions and accents.

Impact on clinical outcomes

- Most peer-reviewed studies of digital scribes evaluated DAX and Nabla. The strongest available evidence is for reduction in documentation burden, derived from a large, controlled study of Nabla (n=303,266 encounters). However, studies are unable to link reduction in documentation time to improvements in clinician workflow and well-being. Evidence concerning physician workload and burnout, and patient experience was more limited and inconsistent.
- Note accuracy remains an important consideration. Studies reported variable accuracy across components of the clinical note, with documented instances of hallucination, and difficulty with highly scripted or complex encounters.

Implementation considerations

Selection of a particular solution at the MUHC needs to consider various factors including:

- EHR integration and editing burden
- Overall note quality
- Customization such as specialty-specific templates
- Bilingual performance
- Data privacy and consent
- Clinical impact, e.g. actual time saved
- Cost impact

CONCLUSIONS

- Ambient AI digital scribes are a **promising but still maturing technology** with the potential to reduce physician documentation burden. The available evidence provides some support for reduced documentation time, but evidence regarding ease of use, accuracy, physician workload and burnout, and patient experience remains sparse or inconsistent. No individual product can currently be identified as clearly superior based on published clinical evidence.
- The market also presents an important trade-off. Larger and more established products such as DAX and Nabla have the strongest published evidence base and more developed EHR integration but are among the more expensive options. Newer Canadian and Quebec-based products may provide advantages in local privacy alignment, cost, and potentially language and accent recognition but generally have less independent clinical evidence and less mature EHR integration.

RECOMMENDATIONS

Given the limited evidence on whether digital AI scribes improve documentation clarity and accuracy while reducing physician workload and burnout, **TAU recommends using implementation as an opportunity to generate real-world evidence through a limited, systematically evaluated pilot before considering broader organizational adoption.**

We recommend:

- A limited pilot in a defined setting to measure:
 - Clinician experience (ease of use) and satisfaction
 - Clinician workflow integration and documentation burden
 - Documentation accuracy and risk of bias
 - Concerns around data security, consent, and potential misuse of sensitive information
 - Patient experience, including consent issues
 - Compatibility with EHR and other software
 - Cost impact;
- Establishing organizational privacy and consent requirements;
- Maintaining clinician accountability for documentation;
- The development of continuous monitoring and benchmarking processes as these tools, as well as the underlying algorithms of large language models, continue to rapidly evolve.

SOMMAIRE

Assistants de rédaction numériques basés sur l'IA ambiante pour la documentation clinique : analyse rapide du paysage et comparaison des fournisseurs

CONTEXTE

Les assistants de documentation numérique (ou « scribes numériques ») basés sur l'IA ambiante utilisent la reconnaissance vocale et l'IA générative pour capter les échanges entre cliniciens et patients et générer des ébauches de documentation clinique. Ces technologies offrent le potentiel d'alléger la charge administrative, d'optimiser les flux de travail cliniques et de permettre aux cliniciens de consacrer davantage d'attention aux patients lors des consultations. Toutefois, le marché évolue rapidement ; il existe des différences notables entre les produits en matière d'intégration aux dossiers de santé numériques (DSN), de fonctionnalités, de confidentialité et d'architecture des données, tandis que les données probantes indépendantes concernant leurs avantages cliniques et opérationnels restent limitées.

OBJECTIF

Comparer des solutions de scribes numériques basées sur l'IA ambiante adaptées au contexte canadien et à celui du CUSM, en évaluant leurs fonctionnalités, leurs avantages cliniques, leurs points forts, leurs limites et les enjeux de confidentialité, afin d'identifier des solutions viables susceptibles d'améliorer l'efficacité de la documentation et l'expérience des cliniciens, tout en respectant les exigences réglementaires et celles du Québec en matière de confidentialité.

MÉTHODES

Analyse de l'environnement

- Nous avons réalisé une analyse rapide de l'environnement portant sur neuf solutions de scribes numériques basées sur l'IA ambiante : DAX Copilot, Nabla, Ambience, Scribeberry, NoteDx, Heidi, Coeurway, Tali AI et Plume AI.
- Les produits ont été comparés selon les critères suivants : fonctionnalités principales, intégration aux DSN et interopérabilité, capacités linguistiques et de personnalisation, confidentialité et hébergement des données, coûts et exigences de mise en œuvre.
- Cette analyse a été complétée par un examen ciblé des données probantes publiées sur les résultats cliniques et opérationnels, ainsi que des retours d'expérience de cliniciens accessibles au public. Les affirmations des fournisseurs concernant les résultats cliniques ont été exclues de l'évaluation de l'efficacité.

PORTÉE ET LIMITES DE L'ANALYSE

Il s'agissait d'une analyse rapide de l'environnement plutôt que d'une revue systématique ou d'une évaluation formelle dans le cadre d'un processus d'acquisition. Les caractéristiques des fournisseurs ont été établies à partir d'informations accessibles au public, n'ont pas fait l'objet d'une vérification indépendante par TAU et sont susceptibles d'évoluer rapidement. Aucune étude comparative directe n'a été recensée ; par conséquent, les conclusions ne doivent pas être interprétées comme établissant la supériorité clinique d'un produit par rapport à un autre.

PRINCIPAUX RÉSULTATS

Fonctionnalités principales

- Les neuf produits offraient tous la fonctionnalité de base permettant de capter les échanges lors des consultations cliniques et de générer une ébauche de documentation clinique.
- Les trois principaux fournisseurs — **Dragon Ambient eXperience (DAX) Co-pilot** de Microsoft, **Nabla** (France) et **Ambience** (États-Unis) — proposaient les solutions les plus complètes en termes de fonctionnalités, notamment l'intégration aux dossiers de santé électroniques (DSN), la possibilité d'extraire des données du DSN pour les consulter dans l'application de transcription et une aide intégrée au codage CIM-10. Les produits plus récents ou issus de plus petites entreprises reposaient davantage sur le copier-coller de la documentation générée par les cliniciens dans le DSN. Scribeberry, un produit canadien déjà utilisé au CUSM, proposait également une fonction d'interaction (« talk-back ») permettant aux cliniciens d'interroger l'assistant au sujet de la consultation.
- Les fournisseurs canadiens peuvent offrir des avantages en matière de mise en œuvre grâce à leur connaissance du contexte local relatif à la confidentialité. L'analyse a permis d'identifier cinq fournisseurs canadiens (**Scribeberry, NoteDx, Tali AI, Coeurway et Plume AI**) affirmant se conformer aux exigences canadiennes et québécoises en matière de protection de la vie privée.
- Les produits développés au Québec (**Coeurway, Plume AI**) pourraient offrir l'avantage pratique supplémentaire de mieux reconnaître les expressions linguistiques et les accents locaux.

Impact sur les résultats cliniques

- La plupart des études évaluées par des pairs portant sur les assistants de documentation numériques (ou « scribes numériques ») ont analysé les solutions DAX et Nabla. Les données probantes les plus solides concernent la réduction de la charge de documentation ; elles proviennent d'une vaste étude contrôlée sur Nabla (n = 303 266 consultations). Toutefois, les études ne parviennent pas à établir un lien direct entre la réduction du temps consacré à la documentation et l'amélioration du flux de travail ou du bien-être des cliniciens. Les données concernant la charge de travail et l'épuisement

professionnel des médecins, ainsi que l'expérience des patients, sont plus limitées et présentent des résultats divergents.

- La précision des notes demeure un facteur important. Les études font état d'une précision variable selon les différentes sections de la note clinique, signalant des cas avérés d'« hallucinations » (génération d'informations erronées) ainsi que des difficultés lors de consultations très structurées ou complexes.

Considérations relatives à la mise en œuvre

Le choix d'une solution spécifique pour le CUSM doit prendre en compte divers facteurs, notamment :

- L'intégration au dossier de santé numériques (DSN) et la charge de révision des notes
- La qualité globale des notes
- Les possibilités de personnalisation (ex. : modèles adaptés à une spécialité)
- Les performances en contexte bilingue
- La confidentialité des données et le consentement
- L'impact clinique (ex. : gain de temps réel)
- L'impact financier

CONCLUSIONS

- Les assistants de documentation numériques basés sur l'IA ambiante constituent une technologie prometteuse, bien qu'encore en phase de maturation, susceptible de réduire la charge de documentation des médecins. Les données disponibles étayent l'idée d'une réduction du temps consacré à la documentation, mais les preuves concernant la facilité d'utilisation, la précision, la charge de travail et l'épuisement professionnel des médecins, ainsi que l'expérience des patients, restent rares ou contradictoires. À l'heure actuelle, aucune solution ne peut être clairement désignée comme supérieure sur la base des données cliniques publiées.
- Le marché présente également un arbitrage important. Les produits les plus vastes et les mieux établis, tels que DAX et Nabla, s'appuient sur les données probantes les plus solides et offrent une intégration plus poussée au DSN, mais ils figurent parmi les options les plus coûteuses. Les produits plus récents, issus du Canada ou du Québec, peuvent offrir des avantages en matière de conformité aux normes locales de confidentialité, de coût et potentiellement de reconnaissance des langues et des accents, mais ils disposent généralement de moins de données cliniques indépendantes et d'une intégration au DSN moins aboutie.

RECOMMANDATIONS

- Compte tenu du peu de données probantes concernant l'efficacité des assistants IA numériques pour améliorer la clarté et la précision de la documentation, tout en réduisant la charge de travail et l'épuisement professionnel des médecins, **TAU recommande de profiter de la mise en œuvre de ces assistants pour générer des données concrètes grâce à un projet pilote limité et systématiquement évalué, avant d'envisager une adoption plus large au sein de l'organisation.**

Nous recommandons :

- Un projet pilote limité, mené dans un contexte défini, afin de mesurer :
 - L'expérience des cliniciens (facilité d'utilisation) et leur satisfaction
 - L'intégration au flux de travail des cliniciens et la charge de travail liée à la documentation
 - La précision de la documentation et le risque de biais
 - Les préoccupations relatives à la sécurité des données, au consentement et à l'utilisation abusive potentielle des informations sensibles
 - L'expérience des patients, notamment en ce qui concerne le consentement
 - La compatibilité avec le dossier médical électronique (DME) et les autres logiciels
 - L'impact sur les coûts ;
- L'établissement d'exigences organisationnelles en matière de confidentialité et de consentement ;
- Le maintien de la responsabilité des cliniciens en matière de documentation ;
- Le développement de processus de surveillance continue et d'évaluation comparative, car ces outils, ainsi que les algorithmes sous-jacents des grands modèles de langage, continuent d'évoluer rapidement.

Ambient AI Digital Scribes for Clinical Documentation: A Rapid Environmental Scan and Vendor Comparison

1. BACKGROUND

Clinical documentation is a substantial component of physician workload. Conventional documentation approaches, including typing and traditional voice-recognition systems, may require clinicians to divide their attention between the patient and the electronic health record (EHR) and can contribute to documentation outside scheduled clinical hours.¹

Ambient AI scribes represent a newer approach to documentation. Rather than requiring clinicians to dictate a note explicitly, these systems can listen to the clinical encounter via a smartphone microphone, identify clinically relevant information, and use large language models (LLMs) to generate a structured draft note for clinician review and approval.¹

The availability of these products has expanded rapidly, including several Canadian and Quebec-based solutions. This creates an opportunity for improved clinical documentation but also raises questions regarding comparative performance, privacy, integration, implementation requirements, and value within the MUHC environment.

2. OBJECTIVE

The evaluation questions (objectives of this report) were the following:

1. How do different commercially available vendors compare to each other in terms of core functionalities such as real-time transcription, EHR integration, usability, and compliance with privacy and regulatory requirements?
2. What is the available evidence on clinical and operational outcomes including documentation burden, note quality, clinical workload and burnout, and patient experience?

3. METHODS

3.1 Rapid environmental scan

We undertook a structured environmental scan to identify major international ambient AI scribe vendors, Canadian and Quebec-based solutions, and products already being used or considered within the MUHC. The vendor scan was supplemented by a targeted review of published literature examining the impact of ambient/digital scribes on documentation

burden, clinician satisfaction and well-being, and patient experience. We also reviewed important organizational, provider, patient, and technology-level implementation considerations.

Information was compiled into a standardized vendor comparison covering:

- **Core functionalities:** real-time documentation and transcription; EHR integration and interoperability; customization; referral letter generation and coding assistance
- **Québec/Canada fit:** French-language capability; privacy, security and data-hosting arrangements
- **Clinical benefit:** available evidence concerning documentation burden, accuracy, clinician and patient experience
- Pricing models and availability in Canada

3.2 Scope and limitations

This was a rapid environmental scan rather than a systematic review or formal procurement evaluation. Vendor characteristics were derived partly from vendor-reported information and may change rapidly. Direct head-to-head comparative studies were not identified; consequently, the findings should not be interpreted as establishing the clinical superiority of one product over another.

4. RESULTS

4.1 Vendor landscape and functionality

We identified nine ambient AI digital scribe products (DAX Copilot, Nabla, Ambience, Scribeberry, NoteDx, Heidi, Coeurway, Tali AI, Plume AI) through the environmental scan, including three products already being used within the MUHC and several Canadian and Quebec-based solutions ([Table 1](#)).

Although all evaluated products provided the basic functions required of an ambient scribe, important differences emerged in the breadth of functionality and degree of integration with existing clinical information systems ([Table 2](#)).

- The 3 largest vendors, **DAX Co-pilot**, **Nabla** and **Ambience** offered the most comprehensive support for EHR integration and integrated ICD-10 coding assistance.
- All Canadian vendors (**Scribeberry**, **NoteDx**, **Tali AI**, **Coeurway**, and **Plume AI**) reported being compliant with Canadian and Quebec privacy requirements.
- Quebec-developed products (**Coeurway**, **Plume AI**) may have the additional practical advantage of better recognizing local language expressions and accents.

Given that product capabilities are evolving rapidly, the presence of individual features should be confirmed at the time of procurement rather than assumed based on this environmental scan.

Table 1. Ambient AI scribe landscape*

Vendor	EHR integration	Published evidence	Canadian/Québec considerations	Notable functionality
DAX Copilot	Direct integration reported	Yes	International vendor	EHR integration, coding, EHR information retrieval, AI-generated referral letters
Nabla	Direct integration reported	Yes	International vendor	EHR integration, coding, EHR information retrieval, AI-generated referral letters
Ambience	Direct integration reported	Limited	International vendor	EHR integration, coding, EHR information retrieval, AI-generated referral letters
Scribeberry	Limited/copy-paste	No independent evidence identified	Canadian	Talk-back functionality, AI-generated referral letters
NoteDx	Limited/copy-paste	No independent evidence identified	Québec/Canadian	AI-generated referral letters, bilingual support
Heidi	Enterprise integration available	No independent evidence identified	International; Canadian use	Custom templates, specialty-specific documentation
Coeurway	Limited/copy-paste	No independent evidence identified	Québec	AI-generated referral letters, Local language and accent recognition
Tali AI	Limited/copy-paste	No independent evidence identified	Canadian	AI-generated referral letters, multilingual capability
Plume AI	Limited/copy-paste	No independent evidence identified	Québec	Local language and accent recognition

**Vendor functionality and geographic/privacy characteristics are based on publicly available information identified during the TAU environmental scan and were not independently verified. Product capabilities may have changed since the time of the scan. "No independent evidence identified" indicates that TAU did not identify vendor-specific peer-reviewed outcome evidence in its search; it does not indicate that the product has no evidence.*

Table 2. Main differences in vendor functionalities

Key functionality	Purpose	Illustrative vendor examples	Why this is important
EHR integration	Support direct EHR integration, retrieval of EHR information for use within the scribe application, and integrated ICD-10 coding assistance.	- DAX Copilot (Microsoft/Nuance), Nabla, and Ambience - Smaller and more recently developed products generally had less extensive EHR integration and frequently required clinicians to copy and paste generated notes into the EHR.	- Potential workflow gains may be diminished if clinicians need to move repeatedly between applications or manually transfer documentation into the EMR. - The degree and quality of EHR integration should therefore be considered a key criterion in vendor selection.
Customization	Custom templates that are specialty- or user-specific	Scribeberry offers a “talk-back” function for users to interact conversationally with the scribe after an encounter	- Users report challenges with highly scripted or complex medical information - Specialty-specific templates, note structure, level of detail and adaptation to individual physician documentation styles may influence adoption.
Quebec/Canada fit			
Privacy and data architecture	Data location, audio retention, third-party cloud and LLM providers, model training practices and compliance with Quebec requirements	Law 25 compliant: Plume IA, Coeurway, Scribeberry, Tali AI, Heidi PIPEDA compliant: Scribeberry, NoteDx, Coeurway, Tali AI, Heidi	Local products may offer advantages in privacy/data residency.
Language	Supports French and regional accents	Plume IA, Coeurway	Recognition of local accents or expressions is an advantage
Evidence maturity	Published vendor-specific evidence was concentrated primarily on DAX Copilot and Nabla. For most other products, evidence for clinical outcomes and note quality was not identified in this scan.	DAX Copilot, Nabla	Establishes value of AI scribes for impact on clinician workload and well-being

4.2 Published Evidence of Clinical and Operational Benefit

Peer-reviewed evidence was concentrated predominantly on **DAX Copilot and Nabla**, which were the most widely studied products identified in the scan (Table 3). Detailed results are shown in Appendix Table A-1.

Table 3. Evidence of clinical benefit

Outcome	Findings	Conclusion
Documentation burden	A large, controlled study of Nabla users (303,266 encounters) reported a statistically significant reduction in documentation time outside work hours, and time spent in notes during appointments, comparing users versus nonusers (Tierney et al.).	Some evidence of reduction
Transcript accuracy	- DAX users noted variable accuracy with certain sections like patient history and the Assessment and Plan sections of the note (Nguyen et al.). - An analysis of 35 Nabla transcripts found excellent quality of transcripts based on attributes including accurate, free from bias, internally consistent, but some instances of hallucinations were noted (Tierney et al.).	Mixed
Patient experience	1 study reported no difference in likelihood to recommend the tool (Haberle et al.). - Another study reported some patients expressed unease at having their visits recorded on a smartphone (Nguyen et al.). - A patient survey of 21 Nabla encounters reported neutral to positive experiences; 71% reported more time spent speaking with the physician (Tierney et al.).	Evidence gap: potential benefit needs to be balanced against privacy/consent concerns
Clinician well-being	2 small studies reported contradicting results on physician burnout. The study reporting positive results was a small (n=38) pre-post physician survey, hence results are exploratory.	Mixed

4.2.1 Documentation burden

- The clearest evidence of benefit concerned documentation burden. A large, controlled evaluation of Nabla involving 3,442 physicians (303,266 encounters) used a difference-in-differences analysis to compare users versus non-users before and after implementation. The study reported a statistically significant reduction for primary care physicians in time spent on note documentation outside of working hours and in time spent in notes during appointments.¹ While this study adjusted for clinician, specialty and appointment volume, there was considerable heterogeneity in the metrics measured, and absolute time saved was difficult to interpret. Therefore, the quality improvement design of this study limits these findings as exploratory.

- Three smaller studies reported modest decreases in documentation time of around 1 minute per patient.²⁻⁴

These findings provide **some evidence that ambient AI scribes can reduce documentation time**. However, the magnitude and consistency of benefit vary between clinicians and clinical contexts, with studies often being unable to link reduced note-writing time to improved overall clinical productivity and physician well-being.

4.2.2 Note accuracy and quality

Accuracy remains a key clinical safety consideration. Users of DAX reported variable accuracy across different sections of the clinical note, including patient history and Assessment and Plan sections.⁵ An evaluation of 35 Nabla-generated transcripts reported high overall quality according to attributes such as accuracy, freedom from bias, and internal consistency; however, instances of hallucination were also identified.¹

4.2.3 Patient experience

Evidence demonstrating improvements in patient experience was not identified. One study found no difference in patients' likelihood of recommending care,³ while another reported that some patients expressed discomfort with having their clinical encounter recorded on a smartphone.⁵ A survey of 21 patients with Nabla experience described neutral to positive experiences with the scribe, with 71% reporting spending more time speaking with their doctor and 81% reporting their physician spent less time looking at the computer screen.¹

4.2.4 Provider experience and well-being

Evidence regarding physician experience and burnout remains limited and inconsistent, with two small studies reporting conflicting findings.^{5,6} The positive study involved a survey of 38 physicians who self-reported on burnout, burden, and usability, and reported a statistically significant reduction on a 4-item physician task-load score of 24 points. The survey was conducted before and after implementation over a 3-month period.⁶ It therefore carries a high risk of bias due to its small, pre-post design without temporal controls; its findings should be interpreted as exploratory.

4.3 Canadian/Quebec Context

Privacy and information governance are essential determinants of whether an ambient AI scribe can be implemented safely within the MUHC.

All five Canadian vendors, **Scribeberry, NoteDx, Tali AI, Coeurway, and Plume AI**, reported compliance with Canada's PIPEDA (Personal Information Protection and Electronic Documents Act) and Quebec's Law 5 governing personal information privacy and security.

Heidi, an Australian vendor already being used within the MUHC, also reported compliance with PIPEDA.

4.4 Cost and Value Considerations

Substantial variation in price was identified across the products reviewed. **DAX and Nabla were among the most expensive solutions**, whereas several smaller vendors offered free or lower-cost subscription options that may be more attractive to individual clinicians.

5. DISCUSSION: IMPLEMENTATION CONSIDERATIONS

Selection of a particular solution at the MUHC should focus on clinical accuracy, editing burden, EHR integration, bilingual performance, usability, privacy, and actual time saved.

5.1 Evidence maturity

The available evidence, mostly derived from evaluations of DAX Copilot and Nabla, is strongest for reductions in documentation time. However, these varied according to clinical specialty and clinician workflow, editing needs and EHR integration.

Importantly, studies were often unable to link reduced note-writing time to improved overall clinical productivity and physician well-being. A technology may save time generating documentation while simultaneously introducing other workflow demands, including reviewing and correcting notes, obtaining patient consent, navigating between applications, or reconciling generated information with the EHR.

This suggests that implementation and training requirements may differ substantially between clinicians and specialties, and local context is important during initial adoption.

5.2 Note accuracy and quality

Published evidence and user reports suggest that ambient scribes may perform less reliably during highly complex or strongly scripted clinical encounters. These findings reinforce the need to evaluate note accuracy within specific clinical settings, and underscores that AI-generated documentation should be considered a draft requiring physician review, correction, and approval and cannot replace clinician judgement.

5.3 Patient experience

Ambient documentation could theoretically improve patient–clinician interaction by reducing clinicians' need to type during an encounter, but this potential benefit needs to be balanced

against concerns regarding recording, privacy, transparency, and consent; these factors could benefit from a local evaluation.

5.4 Canadian/Quebec Context

Canadian and Quebec-based vendors may offer potential implementation advantages. In addition to familiarity with Canadian privacy requirements, Quebec-developed digital scribes may have greater familiarity with local expressions and accents, which is relevant in a bilingual clinical environment.

These characteristics should nevertheless be independently validated. Vendor-reported compliance does not replace an MUHC privacy, information-security, legal, and procurement assessment. Prior to implementation, the organization should establish:

- where audio recordings, transcripts, prompts, and generated notes are processed and stored;
- whether information leaves Quebec or Canada;
- how long source audio and derived information are retained;
- whether identifiable or de-identified clinical information is used to train or improve AI models;
- which third-party cloud, speech-recognition, or large language model providers receive information;
- encryption, authentication, audit, and access-control arrangements;
- deletion and breach-response procedures; and
- whether changes in underlying AI models could alter privacy or security characteristics after implementation.

Patient consent and transparency requirements should also be defined prospectively and implemented consistently across users. Canada Health Infoway has published a helpful [guide](#) to implementing AI scribes in compliance with Quebec privacy requirements.

5.5 Cost and overall value

While the cost of ambient scribe solutions varied widely, price alone is unlikely to provide a meaningful indication of value.

The total cost of adoption should include not just acquisition costs (software licensing and subscription fees) but also:

- EHR integration and interface development;
- implementation and configuration;
- privacy and cybersecurity assessment;
- clinician training and change management;
- technical and clinical support;

- ongoing upgrades and maintenance; and
- clinician time required to review or correct generated documentation.

A more expensive integrated product may potentially produce greater workflow benefits than a lower-cost solution requiring extensive manual transfer of documentation. Conversely, a smaller Canadian product may provide sufficient functionality at substantially lower cost if it performs adequately within the intended clinical workflow.

Local measurement of benefits and costs should therefore inform any subsequent procurement decision.

6. CONCLUSIONS

- Ambient AI digital scribes are a promising but still maturing technology with the potential to reduce physician documentation burden. The available evidence provides some support for reduced documentation time, but evidence regarding ease of use, accuracy, physician workload and burnout, and patient experience remains sparse or inconsistent. No individual product can currently be identified as clearly superior based on published clinical evidence.
- The market also presents an important trade-off. Larger and more established products such as DAX and Nabla have the strongest published evidence base and more developed EHR integration but are among the more expensive options. Newer Canadian and Quebec-based products may provide advantages in local privacy alignment, cost, and potentially language and accent recognition but generally have less independent clinical evidence and less mature EHR integration.

7. RECOMMENDATIONS

Given the limited evidence on whether digital AI scribes improve documentation clarity and accuracy while reducing physician workload and burnout, **TAU recommends using implementation as an opportunity to generate real-world evidence through a limited, systematically evaluated pilot before considering broader adoption.**

We recommend:

- **A limited pilot** in a defined setting to measure:
 - Clinician experience (ease of use) and satisfaction
 - Clinician workflow integration and documentation burden
 - Documentation accuracy and risk of bias

- Concerns around data security, consent, and potential misuse of sensitive information
 - Patient experience, including consent issues
 - Compatibility with EHR and other software
 - Cost impact;
- Establishing organizational privacy and consent requirements;
- Maintaining clinician accountability for documentation
- The development of continuous monitoring and benchmarking processes as these tools, as well as the underlying algorithms of large language models, continue to rapidly evolve.

APPENDICES

APPENDIX A DETAILED RESULTS OF INCLUDED STUDIES

Table A-1. Findings from included studies on clinical outcomes

Title	Setting	Study design	Population	Documentation			Outcomes			Study Conclusion	
				Documentation time	Productivity	Patient satisfaction	Ease of use	Accuracy	Feasibility and Acceptability		Clinical wellbeing
DAX											
Haberle, T., et al., The impact of nuance DAX ambient listening AI documentation: a cohort study. J Am Med Inform Assoc, 2024. 31(4): p. 975-979.	Intermountain Health (Intermountain), a 33-hospital integrated health system, Utah	Prospective controlled study	99 DAX users 99 non-users	A statistically significant decrease in documentation time (min/patient) from 5.3 to 4.5 min/pt.	No difference in productivity (measured as work relative values unit)	No difference in likelihood to recommend				Nuance DAX use showed positive trends in provider engagement at no risk to patient safety, experience, or clinical documentation. There were no significant benefits to patient experience, documentation, or measures of provider productivity.	
Liu T-L, Hetherington Timothy C, Dharod A, et al. Does AI-Powered Clinical Documentation Enhance Clinician Efficiency? A Longitudinal Study. NEJM AI. 2024/11/27 2024;1(12):Aloa2400659.	Atrium Health, large multisite academic centre, NC	Prospective controlled study June - Aug 2023	112 primary care clinicians DAX users 103 non-users	Small decreases in documentation hours could result from high DAX usage (means ratio [MR] 0.93, 95% confidence interval [CI] 0.88 to 0.98)						Article concluded that their findings suggest that the tool did not make clinicians as a group more efficient.	
Oliver T. Nguyen et al., Implementing Digital Scribes to Reduce Electronic Health Record Documentation Burden Among Cancer Care Clinicians: A Mixed-Methods Pilot Study. JCO Clin Cancer Inform 7, e2200166(2023).	National Cancer Institute (NCI)–designated Comprehensive Cancer Center in the south-eastern United States	Survey	21 cancer care clinicians (only 9 completed 1-month survey)	Mixed: some reported no impact on documentation efficiency			Varying amounts of time to feel comfortable with tool	DS helped with some parts of the note (e.g., history), it struggled with other sections (e.g., patient education). Instances of errors were also reported, such as the labeling of speakers. More edits were needed on the Assessment and Plan sections of the note. Fewer edits were needed for the review of systems	A mean score of 16 or higher was indicative that the DS was feasible, acceptable, and appropriate. Although the total feasibility scores were slightly lower than our cutoff point (15.2 v 16.0), clinicians rated the DS as marginally acceptable (16.0) and appropriate (16.3).	This study reported no significant differences in burnout (3.6 v 3.9, P = .081) 1 month after digital scribe use.	Preliminary findings suggest that DS implementation is marginally acceptable, appropriate, and usable among cancer care clinicians. Individualized training and on-site support may improve implementation.

Title	Setting	Study design	Population	Outcomes							Study Conclusion
				Documentation time	Productivity	Patient satisfaction	Ease of use	Accuracy	Feasibility and Acceptability	Clinical wellbeing	
DOI:10.1200/CCO.22.00166								and physical examination results sections. Mixed results were reported for the history sections with some reporting few edits and others reporting many edits.	Usability was considered marginally usable (68.6 v 68.0)		
Stephen P Ma, et al., Ambient artificial intelligence scribes: utilization and impact on documentation time, Journal of the American Medical Informatics Association, Volume 32, Issue 2, February 2025, Pages 381–385, https://doi.org/10.1093/jamia/ocae304	Stanford Health Care	Prospective uncontrolled study October 2023 to January 2024	45 physicians	Time spent on clinical documentation reduced by 0.57 minutes							Large language model-powered ambient AI scribes may reduce documentation burden. Further studies are needed to identify which users benefit most from current technology and how future iterations can support a broader audience.
Shreya J Shah, et al., Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden, Journal of the American Medical Informatics Association, Volume 32, Issue 2, February 2025, Pages 375–380, https://doi.org/10.1093/jamia/ocae295	Stanford Health Care	Paired pre- and post-Survey analysis October 2023 to January 2025	38 physicians					Physicians' perception of planning to use DAX Copilot often when completing documentation significantly decreased in the post-survey period.	Physicians' We observed a large statistically significant reduction (–24.42 on 0–100 scale; p < .001) in the 4-item physician task-load score derivative and burnout (–1.94 on 0–10 scale; p < .001) assessed via the 4-item work exhaustion score		Ambient AI scribes like DAX Copilot may enhance clinical workflows. Further research is needed to optimize widespread implementation and evaluate long-term impacts.
NABLA											

Title	Setting	Study design	Population	Outcomes							Study Conclusion
				Documentation time	Productivity	Patient satisfaction	Ease of use	Accuracy	Feasibility and Acceptability	Clinical wellbeing	
Tierney Aaron A, Gayre G, Hoberman B, et al. Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation. NEJM Catalyst;5:CAT.23.0404.	The Permanente Medical Group (TPMG), a multidisciplinary physician group including >9,000 physicians working in the Kaiser Permanente Northern California	Prospective controlled before-after study October 16, 2023 to December 24, 2023	3,442 physicians and staff across diverse clinical specialties	There was a statistically significant reduction in time in notes per appointment, comparing users (before and after) versus nonusers (before and after) with difference-in-differences (DiD) analysis.				Overall, the transcripts averaged a score of 48 out of a possible 50 points with the highest average ratings (average value, >4.95) achieved in the free from bias, synthesized, internally consistent, and succinct domains. Modestly lower, but still considered excellent, average ratings (average, 4.6 to 4.7) were noted in the thorough, organized, and accurate domains. There were a few instances of hallucination noted			-Ambient artificial intelligence (AI) scribes show early promise in reducing clinicians' burden; - Ambient AI scribes were found to be acceptable among clinicians and patients; - Although a review of 35 AI-generated transcripts resulted in an average score of 48 of 50 in 10 key domains, AI scribes are not a replacement for clinicians; - Given the incredible pace of change, building a dynamic evaluation framework is essential to assess the performance of AI scribes across domains including engagement, effectiveness, quality, and safety

REFERENCES

1. Tierney A, Gayre G, Hoberman B, et al. Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation. *NEJM Catalyst Innovations in Care Delivery*. 2024/02/21 2024;5(3).
2. Ma SP, Liang AS, Shah SJ, et al. Ambient artificial intelligence scribes: utilization and impact on documentation time. *J Am Med Inform Assoc*. Feb 1 2025;32(2):381–385.
3. Haberle T, Cleveland C, Snow GL, et al. The impact of nuance DAX ambient listening AI documentation: a cohort study. *J Am Med Inform Assoc*. Apr 3 2024;31(4):975–979.
4. Liu T-L, Hetherington T, Dharod A, et al. Does AI-Powered Clinical Documentation Enhance Clinician Efficiency? A Longitudinal Study. *NEJM AI*. 2024/11/27 2024;1(12).
5. Nguyen OT, Turner K, Charles D, et al. Implementing Digital Scribes to Reduce Electronic Health Record Documentation Burden Among Cancer Care Clinicians: A Mixed-Methods Pilot Study. *JCO Clin Cancer Inform*. Mar 2023;7:e2200166.
6. Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. *J Am Med Inform Assoc*. Feb 1 2025;32(2):375–380.