Impacts of AI Scribes on Clinical Outcomes, Efficiency, and Documentation

A rapid review

Prepared for Canada Health Infoway

Date : 29-01-2025

Prepared by :

- **Maxime Sasseville; PhD**
- **Farzaneh Yousefi; MSc**
- **Steven Ouellet; PhD**
- **Théo Stefan; MSc**
- **Valérie Carnovale; PhD**
- **Frédéric Bergeron; MSI**
- **Annie LeBlanc; PhD**

To the attention of :
Canada Health Infoway

Suggested citation :
Sasseville, M., Yousefi, F., Ouellet, S., Stefan, T., Carnovale, V., Bergeron, F., LeBlanc, A., (2024) Impacts of AI Scribes on Clinical Outcomes, Efficiency, and Documentation, SPOR Evidence Alliance.

1 Acknowledgements

The research team would like to acknowledge the contribution of the team at Canada Health Infoway, Simon Hagens, Bobby Gheorghiu & Linda Ling.

2 Financing

The knowledge synthesis included in this report was funded by the SPOR Evidence Alliance. The work was carried out under the direction of Prof. Maxime Sasseville and Prof. Annie LeBlanc.

The SPOR Evidence Alliance network is supported by the Canadian Institutes of Health Research (CIHR) under the initiative of Canada's Strategy on Patient-Oriented Research (SPOR).

2.1 Conflicts of interest

The authors of this report declare that they have no conflict of interest in relation to the funding of this review.

3 Abbreviations

SPOR: Strategy for patient-oriented research
AI: artificial intelligence
NLP: Natural language processing
EHR: Electronic Health Records

4 Table of contents

1 Acknowledgements II

2 Financing II

2.1 Conflicts of interest II

3 Abbreviations II

4 Table of contents III

5 Problem statement 5

6 Objective 6

6.1 PICOS elements 6

7 Methods 6

7.1 Research question 6

7.2 Inclusion and exclusion criteria 7

7.3 Bibliographic search strategies for publications 7

7.4 Selection of publications and data extraction 7

7.5 Data analysis strategy 7

7.6 Quality analysis 8

7.7 Data management 8

7.8 Involvement of knowledge users 8

8 Results 9

8.1 Description of studies 9

8.2 Characteristics of the studies 10

8.3 Clinician outcomes 22

8.3.1 Reduced documentation burden 22

8.3.2 Cognitive load 22

8.3.3 Improved provider well-being 22

8.3.4 Mixed impact on productivity 22

8.4 Implementation outcomes 23

8.4.1 Implementation and training needs 23

8.4.2 Patient perspectives and ethical considerations 23

8.5 Accuracy 23

8.5.1 Overall accuracy of AI Scribes 23
8.5.2 Characteristics influencing accuracy: 23
8.6 Results visualization 25
9 Discussion and conclusions 26
9.1.1 Summary of results 26
9.1.2 Strengths of this study and Impacts of the results 26
9.1.3 Limitations 27
9.1.4 Conclusions 27
10 References 28
11 APPENDIX 1 - Strategies for identifying data sources 29

5 Problem statement

Physician burnout is a major issue in the healthcare system, and one significant contributor is the burden of clinical documentation, especially with electronic health records (EHRs) [1, 2]. The demands of documentation take time away from patient care and contribute to physician stress and burnout. Studies have shown that physicians often spend an excessive amount of time on documentation, leading to decreased job satisfaction and a negative impact on work-life balance [2, 3]. The frustration stemming from the documentation burden, coupled with the emotional exhaustion and reduced sense of accomplishment associated with burnout, has significant consequences for physician well-being and patient care.

AI scribes can address the significant challenge of physician burnout, coming in part from excessive time spent on clinical documentation [3, 4]. This burden has negative implications for both physicians and patients, contributing to job dissatisfaction, decreased productivity, and potentially compromised patient care. While human scribes have been employed to alleviate these issues, they come with limitations such as cost, training needs, and high turnover rates. The sources posit that AI scribes may offer a more scalable and sustainable solution by automating documentation tasks and reducing reliance on human scribes.

AI scribes leverage advancements in natural language processing (NLP) and machine learning to record, transcribe, and summarize clinical encounters [5]. They aim to streamline documentation workflows, enhance the quality and accuracy of medical records, and ultimately free up physicians to spend more time interacting with patients.

The accelerating rate of NLP research results in a wide range of AI scribe systems, including both commercially available products and research prototypes, and employs various methodologies to assess their effectiveness and impact on clinical practice. It is thus urgent to synthesize the knowledge on AI scribes impacts but also characteristics of the technology and their implementation in the workflow.

6 Objective

This review aims to evaluate the effectiveness of AI scribes designed to streamline clinical documentation, focusing on outcomes related to clinician well-being, efficiency, and patient engagement.

6.1 PICOS elements

Table 1. Elements of the PICOS model [6]

PICO Elements	Elements in this project
Participants	All healthcare providers engaged in clinical documentation
Intervention	AI tools designed to streamline clinical documentation, including: • Real-time transcription and note-taking • Automated data entry into EHRs • Natural Language Processing (NLP) for clinical summarization, contextual understanding, and structured documentation • Tools transforming spoken interactions into structured clinical notes
Comparison	Usual administrative practices with no additional AI support.
Outcome	• **Clinician outcomes:** stress, burnout, administrative time reduction (during and after clinical encounter), patient engagement focus, cognitive load. • **Efficiency outcomes:** healthcare system efficiency metrics (e.g., wait times, patient throughput, costs). • **Documentation outcomes:** accuracy, relevance, reduction in manual note revisions.
Setting	All clinical settings
Study specifications	• Eligible studies: Randomized controlled trials (RCTs), quasi-experimental designs, prospective cohort studies, pre-post studies, observational, and mixed-method studies. • Excluded studies: Protocols, ongoing studies, qualitative-only studies, reviews, commentaries, and non-AI tools
Language	No restrictions
Year of publication	No restrictions

7 Methods

We followed the Cochrane handbook for systematic reviews for methods guidance[6] with specific adjustments for rapid reviews by a Cochrane methods group [7].

7.1 Research question

What are the impacts of AI scribes on clinical documentation efficiency, administrative burden, clinician stress or burnout, and healthcare system efficiency?

7.2 Inclusion and exclusion criteria

This study focuses on evaluating the impact of AI tools designed to streamline clinical documentation on healthcare providers across all clinical settings. We included interventions based on AI, such as real-time transcription, automated EHR data entry, natural language processing-based clinical summarization, and tools that transform spoken interactions into structured clinical notes. We included all sources that assess the following outcomes: clinician well-being (stress, burnout, cognitive load), patient engagement, healthcare system efficiency (e.g., wait times, costs), and documentation quality (accuracy, relevance, reduced manual revisions). We included all intervention study designs: RCTs, quasi-experimental designs, prospective cohorts, pre-post, observational, and mixed-method studies. We did not limit for language or publication year. We excluded protocols, ongoing studies, qualitative-only studies, reviews, and non-AI interventions.

7.3 Bibliographic search strategies for publications

An experienced librarian (FB) has conducted searches in MEDLINE, Embase, CINAHL, Web of Science, Arxiv, and Google Scholar. We tested the sensitivity the search strategy on three sources provided by our partners at Canada Health Infoway.

7.4 Selection of publications and data extraction

Two reviewers (FY and SO) independently conducted the study selection process. The screening criteria were pilot tested on 10 sources, during which no discrepancies arose. Any decisional conflicts were resolved through discussion or, if necessary, by consulting a third experienced reviewer (MS). Data extraction was performed by one reviewer (FY or SO) using a standardized, pilot-tested form, and the extracted information was validated by an experience reviewer using the full-text source. Discrepancies were again resolved by discussion or, when needed, by a senior third reviewer. Extracted data included study design, population characteristics, intervention specifics, comparator details, outcomes, and key findings.

7.5 Data analysis strategy

We present descriptive data of the included studies. For qualitative data, a content analysis approach was applied, grouping information into themes. Data is summarized narratively. The synthesis focused on providing knowledge users with insights regarding the impact, best practices, gaps, and challenges associated with AI scribes.

7.6 Quality analysis

The quality of the included reviews will be assessed using the Mixed Methods Appraisal Tool [8] by two reviewers. A senior reviewer will verify all scores attributed to the studies. This information will only be presented in the peer-reviewed publication that planned to be submitted for publication in January 2025.

7.7 Data management

We used the Covidence online software [9] to manage the workflow of this review and an Excel sheet to manage the data extraction.

7.8 Involvement of knowledge users

Canada Health Infoway representatives have actively participated in all stages of this rapid review. They were involved in defining the research question, objective, and methods. They reviewed and provided feedback on the work plan before the screening began and have submitted comments on this final report. The representatives will be co-authors in the peer review publication or will be acknowledged if they prefer.

8 Results

8.1 Description of studies

After removing 362 duplicates (344 by Covidence and 18 manually), two reviewers (FY and SO) independently screened 406 titles and abstracts. Following exclusions, 20 references remained for full-text review. Of these, 10 were excluded, leaving 10 references suitable for inclusion in this review. The PRISMA flow diagram [6] of the study inclusion process can be consulted in **Figure 1**.

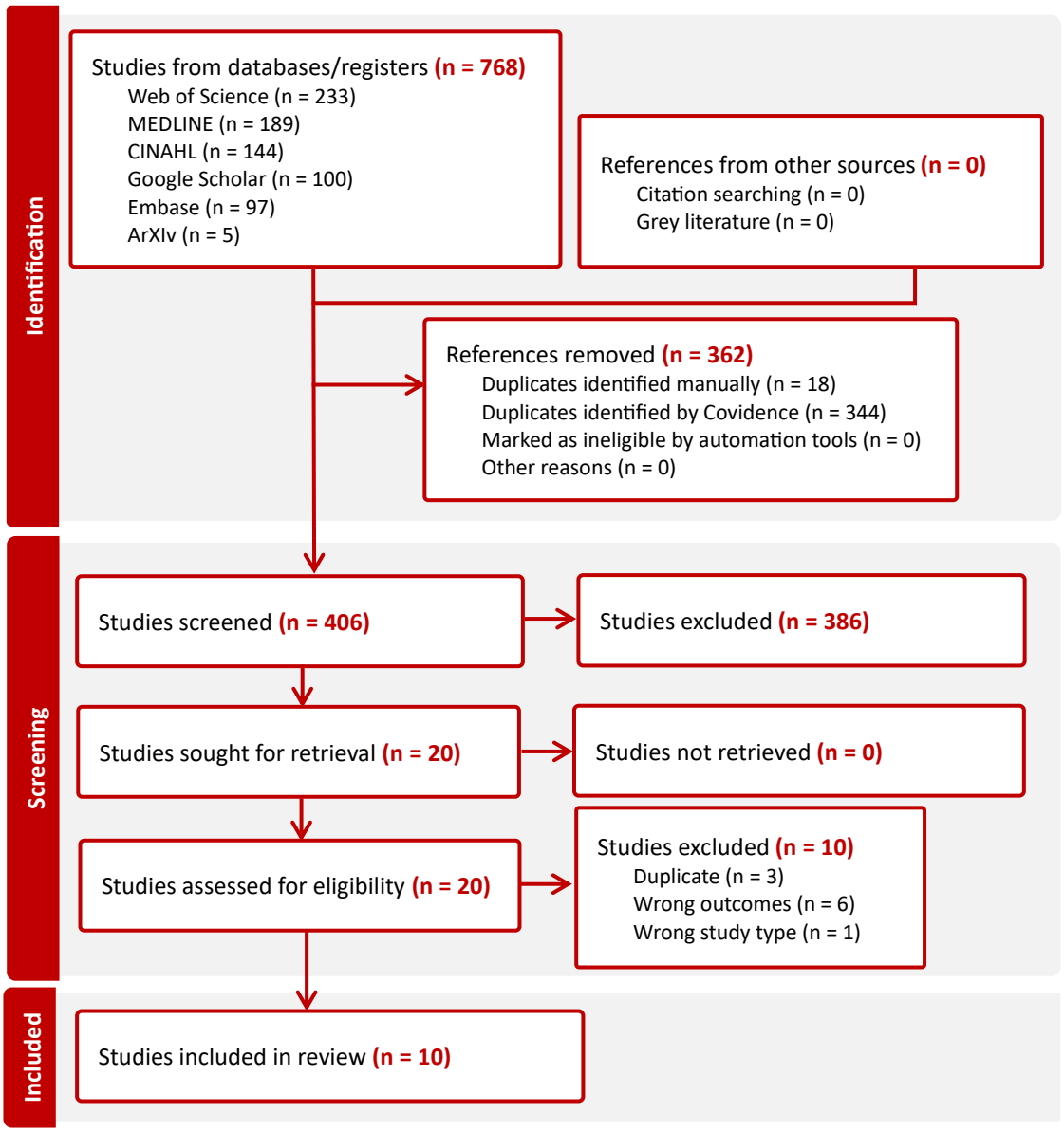

Figure 1. PRISMA flow diagram

8.2 Characteristics of the studies

As shown in **Table 2**, nearly all the included studies were conducted in the United States (8 out of 10), 7 were published in 2024, 2 in 2023, and 1 in 2021.

We identified a diverse range of methodological approaches addressing various aspects of the impacts of AI scribes. A peer-matched controlled cohort study (*Haberle*) enabled a comparative analysis of outcomes, such as caregiver engagement and productivity, involving 99 providers and 76 matched control providers [1]. Two studies utilized simulation-based methodologies: one (*Hata*) combined multi-patient simulations with a quantitative survey to assess AI scribes note accuracy, cognitive and physical load, and provider wellness [2], while the other (*Wang*) focused on the efficiency, training requirements, documentation speed, and reliability of a developed AI scribe using patient-centred communication elements [10].

The development and evaluation of AI scribes were highlighted in a study that focused on the development process (*Islam*) [4]. This study integrated a post-test questionnaire to assess the system's performance, ensuring it met clinicians' needs and enhanced documentation processes. Two comparative studies (*Kernberg, Lee*) [5,6] provided insights into the accuracy and quality of OpenAI's ChatGPT-4 (*Kernberg*) [5] and the performance of GPT-4o Mini compared with Sporo Health's AI Scribe (*Lee*) [6].

Mixed-method approaches were also used. One mixed-method pilot study (*Hudelson*) aimed to explore clinicians' documentation burden and overall experience with AI scribes, combining qualitative and quantitative perspectives [3]. Another mixed-methods longitudinal pilot study (*Nguyen*) assessed changes in clinicians' well-being and documentation burden over time [7]. Lastly, two usability studies employed quantitative descriptive designs (*Sezgin, van Buchem*) [8,9] to assess the performance of AI scribes in supporting clinical documentation, focusing on accuracy (*Sezgin*) [8], efficiency (*van Buchem*) [9], and quality (*van Buchem*) [9].

Table 2. Study characteristics

Studies	Country	Year	Aim of Study	Study Design	Study Setting and Participants	Outcomes Assessed
Haberle et al. *[10]*	USA	2024	To assess the impact of an ambient listening and digital scribing solution, Nuance Dragon Ambient eXperience (DAX), on caregiver engagement, time spent on Electronic Health Records (EHR), including after-hours use, productivity, panel size for value-based care providers, documentation timeliness, and Current Procedural Terminology (CPT) submissions.	Peer-matched controlled cohort study	Outpatient clinics in an integrated healthcare system. 99 providers representing 12 specialties. 76 matched control group providers were included for analysis.	Primary: provider engagement, productivity, panel size, documentation, and coding timeliness. Secondary: patient safety, likelihood to recommend, and number of patients opting out.
Hata et al. *[11]**(Abstract only)	USA	2024	To evaluate the accuracy of an AI scribe and its impact on provider cognitive load, physical load, and wellness.	Multi-patient simulation. Quantitative survey	A simulated Emergency Department (ED) environment. Emergency medicine providers.	Primary: AI scribe's note accuracy, cognitive and physical load, and provider wellness. Secondary: none.
Hudelson et al. *[12]*	USA	2024	1) To identify and test virtual scribe solutions, both live and asynchronous, tailored to the healthcare system's needs; 2) to evaluate and implement these technologies to reduce clinicians' documentation burden, a major contributor to physician burnout.	Mixed methods pilot study	An integrated academic healthcare system. 16 clinicians from diverse specialties.	Primary: clinicians' documentation burden, clinicians' overall experience. Secondary: none.
Islam et al. *[13]*	Bangladesh	2024	1) To develop an automated scribe and intelligent prescribing system for health professionals by identifying user requirements; 2) to design a system that generates medical notes and prescriptions efficiently from voice commands, enhancing the usability of digital scribe solutions, and 3) to evaluate the system's performance to ensure it meets clinicians' needs and improves documentation processes.	AI system development process. Post-test questionnaire	A Medical College Hospital. Enlisted the participation of 17 diabetes patients and 6 doctors.	Primary: similarity rates between AI scribes and prescriptions compared to those generated manually, the system's usability. Secondary: none.
Kernberg et al. *[14]*	USA	2024	To evaluate the accuracy and quality of Subjective, Objective, Assessment, and Plan (SOAP) notes generated by ChatGPT-4 using established History and Physical Examination transcripts as the gold standard, identifying errors and assessing performance across categories.	Comparative Study	14 simulated patient-provider encounters, including professional standardized patients, represented a wide range of ambulatory specialties, and 2 clinical experts.	Primary: an AI model's (*ChatGPT-4*) performance evaluation (e.g., variations in errors, accuracy, and quality of notes generated) using established transcripts of "*History and Physical Examination*" as the gold standard. Secondary: none.
Lee et al. *[15]* *(Preprint)	USA	2024	To evaluate Sporo Health's AI Scribe, a multi-agent system using fine-tuned medical LLMs, by comparing its performance to OpenAI's GPT-4o mini across multiple metrics.	Comparative Study	A dataset of de-identified patient conversation transcripts, and AI-generated summaries, rated by both a medical student and a physician.	Primary: performance evaluation on multiple metrics on clinicians' workflow (e.g., results' reliability, documentation notes quality and billing accuracy) between two (Sporo AI Scribe and GPT-4o mini) multi-agent systems leveraging

						fine-tuned medical LLMs. Secondary: none.
Nguyen et al. *[16]*	USA	2023	To pilot a digital scribe in live clinic settings at a National Cancer Institute–designated Comprehensive Cancer Center, assess its impact on clinician well-being and documentation burden, and identify facilitators and barriers to effective implementation.	Mixed-methods longitudinal pilot study	Clinic settings at a National Cancer Institute–designated Comprehensive Cancer Center, evaluated by 21 "*clinician champions*".	Primary: Impact on clinician well-being and documentation burden, implementation facilitators and barriers for effective AI scribe use, feasibility, and usability. Secondary: Clinicians reported some patients expressed unease at having their visits recorded on a smartphone.
Sezgin et al. *[17]*	USA	2024	To present a proof-of-concept digital scribe system for summarizing Emergency Department consultation calls to support clinical documentation and report its performance.	Usability Study. Quantitative descriptive.	Nationwide Children's Hospital Physician Consult and Transfer Center. 100 phone call recordings from 100 unique callers (physicians) for ED referrals are used.	Primary: Performance (e.g., accuracy rates in medical records, ability to comprehend and replicate the structure and flow of clinical dialogue) of four pre-trained large language models (T5-small39, T5-base39, PEGASUSPubMed47, and BART-Large-CNN46) to support clinical documentation. Secondary: None.
van Buchem et al. *[3]*	Netherlands	2023	To assess the impact of a Dutch digital scribe system on clinical documentation efficiency and quality.	Usability Study	Leiden University Medical Center. 22 medical students with experience in clinical practice and documentation.	Primary: Clinical documentation efficiency (e.g., summarization time) and quality (e.g., accuracy, usefulness). Secondary: None.
Wang et al. *[18]*	USA	2021	To develop a digital scribe for automatic medical documentation using patient-centred communication elements.	Simulation of patient encounters. Quantitative descriptive.	Across multiple departments within a university medical centre with two medical students.	Primary: efficiency, training required, documentation speed, patient-centred communication, and reliability. Secondary: None.

The clinical settings across the 10 included studies reflect a broad spectrum of environments, emphasizing both real-world and simulated healthcare contexts. Outpatient and ambulatory care settings were identified in 2 studies (*Haberle, Nguyen*) [1,7]. Haberle et al. examined outpatient clinics within an integrated healthcare system [1], while Nguyen focused on clinic settings at a National Cancer Institute–designated Comprehensive Cancer Center [7].

Simulated environments were also represented in 4 studies [2,5,6,9]. *Lee et al.* analyzed a dataset of de-identified patient conversation transcripts and AI-generated summaries, evaluated by both a medical student and a physician [6]. *Hata et al.* employed a simulated Emergency Department (ED) environment with emergency medicine providers [2], while *Kernberg et al.* featured simulated patient-provider encounters spanning various ambulatory specialties [5]. To assess the impact of a Dutch digital scribe system on clinical documentation efficiency and quality, *van Buchem et al.* incorporated medical students in a simulated setting at Leiden University Medical Center [9].

Academic and integrated healthcare systems were identified in 3 other studies (*Hudelson, Islam, Wang*) [3,4,10]. *Hudelson et al.* explored a diverse group of 16 clinicians within an integrated academic healthcare system [3], while *Islam et al.* conducted their study in a medical college hospital [4]. *Wang et al.* extended the scope to multiple departments within a university medical centre, engaging two medical students in their analysis [10].

Finally, a specialized context was identified, focusing on phone call recordings from physicians for emergency department referrals at the Nationwide Children's Hospital Physician Consult and Transfer Center (*Sezgin et al.*) [8].

As shown in **Table 3**, the AI scribes represented in the studies demonstrate a wide range of functionalities and designs, each tailored to enhance clinical documentation processes.

Table 3. Intervention (AI-Scribe) characteristics

Studies	Type of AI Scribe (Name)	Key Global Characteristics	How the System Works					
			CD	RT	AN	ADE	AS	AA
Haberle et al. *[10]*	Mobile App (Nuance DAX)	An AI-powered, voice-enabled solution that automatically documents clinical encounters using ambient listening and conversational AI to generate comprehensive documentation from patient-provider conversations.	X	X	X	X	X	X
Hata et al. *[11]*	Software (Not disclosed)	Simply mentioned an ambient artificial intelligence (AI) scribe software	X		X			
Hudelson et al. *[12]*	Mobile App (Not disclosed)	Two virtual scribe solutions were compared: 1) Live Virtual Scribe (All Human-Driven), and 2) Asynchronous Virtual Scribe (Hybrid AI/Human), which used audio recordings, machine learning, and NLP to pre-populate notes, then reviewed and finalized by a human scribe within 4 hours, working asynchronously.	X	X	X	X		
Islam et al. *[13]*	Software (Not disclosed)	The system converts patient voice descriptions into text to generate medical notes and uses extracted medical terms for this purpose. It also creates e-prescriptions from doctors' voice commands. Using NLP and machine learning, it records medical information and generates prescriptions based on voice input from healthcare professionals.	X	X	X	X	X	X
Kernberg et al. *[14]*	Chatbot (Not disclosed)	The ChatGPT-4–generated Subjective, Objective, Assessment, and Plan (SOAP) format is a standard clinical documentation model that organizes interview data into structured headers. It provides a clear framework for healthcare professionals to record and share patient information.	X		X		X	
Lee et al. *[15]*	Device (Sporo and GPT-4o)	Sporo Health's proprietary speaker labelling and transcription AI, and GPT-4o mini hosted within Azure Playground. Fine-tuned AI models using natural language processing (NLP) designed specifically for use in AI scribing.		X			X	
Nguyen et al. *[16]*	Mobile App (Dragon Ambient eXperience)	DS smartphone app: the digital scribe's AI components structured the recorded information into a visit note, and the vendor's staff performed initial editing before the notes were released to the clinician.	X		X			
Sezgin et al. *[17]*	Pre-trained large language models (Not disclosed)	The system converts audio recordings into text: AWS Transcribe transcribes the audio, An annotator reviews and corrects the transcript. Transcription documents are organized as text input for the model, with nurse summary notes used as reference. Four pre-trained language models (*T5-small, T5-base, PEGASUS-PubMed, and BART-Large-CNN*) are employed to summarize clinical conversations based on their strengths in the healthcare domain.					X	
van Buchem et al. *[3]*	Software (Autoscriber)	A web-based tool that transcribes and summarizes medical conversations in Dutch, English, and German. It uses a transformer-based speech-to-text model fine-tuned on clinical data, along with large language models like GPT-3.5 and GPT-4, and a custom prompt structure for summarization. The tool also features self-learning functionality, which was not evaluated in this study.		X			X	
Wang et al. *[18]*	Mobile App/ Software (Not disclose)	It uses automatic speech recognition and natural language processing to transcribe and summarize conversations between healthcare providers and patients into written text.	X	X	X	X	X	

Legends: CD= Clinical Documentation; RT= Real-time Transcription; AN= Automated Notetaking; ADE= Automated data entry; AS= Automated Summarization; AA= Administrative Assistance

Several solutions focus on ambient and automated transcription technologies. *Haberle et al.* [1] highlight Nuance DAX, a voice-enabled, AI-powered solution that uses conversational AI to capture and document clinical encounters through ambient listening. Similarly, *Hata et al.* [2] mention a commercially available ambient AI scribe, though specifics are not disclosed, emphasizing its use in streamlining documentation.

Some studies explore hybrid models that blend AI and human expertise. For example, *Hudelson et al.* [3] contrast two virtual scribe solutions: a fully human-driven Live Virtual Scribe and a hybrid Asynchronous Virtual Scribe, which leverages machine learning and NLP to draft notes from audio recordings that are later refined by human scribes. In another example, *Nguyen et al.* [7] describe the Dragon Ambient eXperience (DAX) smartphone app, which uses AI components to structure recorded information into visit notes, with vendor staff performing initial edits.

We identified 2 studies focusing on the development and application of AI scribe systems [4,5]. *Islam et al.* [4] introduce an automated digital scribe and intelligent prescribing system capable of converting patient voice inputs into text for medical notes and e-prescriptions. Using NLP and machine learning, it extracts medical terms and integrates prescribing capabilities. *Kernberg et al.* [5] explore the use of ChatGPT-4 to generate Subjective, Objective, Assessment, and Plan (SOAP) notes, employing a standardized clinical documentation format to organize patient information.

Advanced language models and speech-to-text systems are also leveraged. *Lee et al.* [6] describe proprietary speaker-labelling transcription AI alongside fine-tuned GPT-4 models, while *Sezgin et al.* [8] use a two-step transcription process with AWS Transcribe and various pre-trained language models (e.g., T5, PEGASUS, BART) to summarize clinical conversations. Another study details Autoscriber, a multilingual web-based tool that employs transformer-based models and large language models (e.g., GPT-3.5, GPT-4) for the transcription and summarization, with a unique self-learning capability [9].

Lastly, *Wang et al.* [10] emphasize a focus on reducing clinician burden. The study describes a patient-centred digital scribe designed to automate medical documentation through automatic speech recognition and NLP. This system prioritizes patient-centred communication while aiming to alleviate physician burnout by efficiently documenting patient interactions.

Regarding how these AI scribe systems work, we also identified in **Table 3** six of the most common features: clinical documentation was reported in 7 studies, automated notetaking in 7, automated summarization in 7, real-time transcription in 6, automated data entry in 4, and administrative assistance in 2 studies (out of 10).

As shown in **Table 4**, we identified the impacts (categorized as positive, uncertain, or negative) for each primary outcome (clinicians, efficiency, and documentation), as well as for patient outcomes (quality of care and satisfaction).

Table 4. Impacts of using AI scribe for each category of outcome.

Studies	Direction of impacts	Clinicians Outcomes	Efficiency Outcomes	Documentation Outcomes	Patient Outcomes
Haberle et al. *[10]*	*Uncertain*	+ Provider engagement in digital scribe • No benefit in productivity for fee-for-service providers • No improvement in total panel size	+ Improving documentation completion + Decreasing time spent in EHR + Decreasing documentation time per patient	+ Decreasing the 24-hour documentation deficiency rate + Increasing 24-hour current procedural terminology submission deficiency rate	• No patient safety events related to the digital scribe • No risk or benefit to patient experience, • No risk or benefit to patient clinical documentation.
Hata et al. *[11]*	*Uncertain/ Positive*	+ Improving provider wellness + Spending more time engaging with patient + Decreasing mental demand + Decreasing physical demand + Decreasing temporal demand + Decreasing frustration + Increasing providers' performance	+ Improving overall efficiency in clinical and non-clinical providers' work + Spending less overall time on providers' work + Freeing up time away from work	- Limitations in the accuracy of notes generated by AI scribes in a complex, multi-patient environment	
Hudelson et al. *[12]*	*Uncertain/ Positive*	+ Reducing documentation burden for clinicians + Improving clinicians' overall experience			
Islam et al. *[13]*	*Positive*		+ High system usability + Effectivity in replicating human-generated documentation + Providing a user-friendly, easy-to-use, intuitive, and interactive interface	+ High accuracy in AI scribe notes	
Kernberg et al. *[14]*	*Negative / Positive*	- Difficult prediction of where errors may occur for healthcare providers	- Negating timesaving by the need for correction - A level of unpredictability, impacting clinical oversight	+ Not making the same errors repeatedly - Inverse correlation between note accuracy and transcript length as well as the amount of reportable data - High error rate, indicating low-quality notes. - The absence of real-time feedback within the application	

				- Limitations in adjusting the models' responses based on user input - Not ability to learn from real-world interactions and refine its output	
Lee et al. *[15]*	*Positive*	+ Clinician satisfaction aligning with the clinician's ground truth while maintaining clarity and relevance with minimal hallucinations	+ Equipped to enhance clinician workflows + Enhancing documentation + the highest standards of privacy and security + High performance in capturing relevant information from the transcript, which may fail by physician-generated note	+ Delivering reliable results despite occasional gaps + Reducing billing inaccuracies + Producing high-quality notes that meet both technical and practical needs + surpass human clinicians in capturing relevant information + reliable tool for clinical documentation	
Nguyen et al. *[16]*	*Negative/ Positive/ Uncertain*	+ Improving clinicians' perceptions of having sufficient documentation time + Improving documentation burden + Reducing editing workload - Not improving the burnout of providers - No impact of DS in helping providers with some parts of the note (e.g. history)	• Varying amounts of time needed to be comfortable using the DS • Improving the editing workload as time progressed • No significant differences in the perceived sufficiency of documentation time • Improving documentation efficiency	- Instances of errors, such as the labelling of speakers - Needing to edit the scribed notes before signing - catching details the clinician may have forgotten to include in the note	- Struggling with other sections (e.g., patient education) - Expressing unease at having patients' visits recorded on a smartphone, at first + Acceptable, appropriate, and usable for patients
Sezgin et al. *[17]*	*Positive/ Negative*	+ reduce the documentation burden on HCPs + effectively complement health care providers' condition tasks	+ A greater ability to comprehend and replicate the structure and flow of clinical dialogue in medical conversation - Inconsistency in performance	+ Providing clean transcripts + High accuracy of replicated notes + The promising accuracy rates - Discrepancy between audio summaries and intake notes - Impacts of external factors like background noise and coughing during conversations	
van Buchem et al. *[3]*	*Negative/ Positive*	+ Good experience with using the tool + Alleviating the administrative burdens faced by healthcare professionals	- Increasing the time spent reading due to the summary length - Increasing analyzing summaries downstream in the clinical process due to the summary length	- The higher word count and lower lexical diversity in the automatic summaries - Providing length summaries - Providing the structure did not always align with their preferences	

			+ Decreasing the time spent on editing the automatic summaries + Improving the quality of summarizing Addressing the challenges of clinical documentation	+ The ability to add context and details, while removing overly general statements and irrelevant text	
Wang et al. *[18]*	*Positive*	+ Needing minimal training to effectively use the digital scribe	+ Efficient documentation + Improving documentation speed Faster note completion notes	+ More reliability of digital scribe	+ Facilitating effective patient-centered communication

Legends: +: Positive Outcome; -: Negative Outcome; •: Uncertain Outcome

Across the 10 included studies, we identified 28 documentation outcomes, 27 efficiency outcomes, 24 clinician outcomes, and 7 patient outcomes. Positive effects of AI scribes were reported in 50% (14/28) of documentation outcomes, 67% (18/27) of efficiency outcomes, 79% (19/24) of clinician outcomes, and 29% (2/7) of patient outcomes.

We also extracted the requirements for implementing an AI-scribe system. These requirements are detailed in Table 5. We identified six main themes. Training support needs and considerations were reported in 6 studies [1,3,7-10], organizational preparation in 5 studies [3,4,7,8,10], technical improvements and conducting evaluations in 4 studies each [1,7,8,10] and [1,3,5,9], respectively, and ethical considerations in 3 studies [7,8,10]. Finally, items for further investigation was proposed in 5 studies [1,4, 8-10].

Table 5.Requirements of Implementing AI- scribes

Main Themes	Items
Further Investigation	• Evaluating the impact of more advanced technology (DAX Co-Pilot) [10] • Comparing the impact of personalization and efficiency training vs Ambient AI documentation [10] • Conducting qualitative research on the benefiting provider and highlighting their needs and expectations [3, 10, 13] • Conducting pilot testing in a controlled environment before full implementation [17, 18] • Evaluating the real-world applicability of digital scribes, including their impact on clinical workflow, quality of care, and patient satisfaction [3]
Technical Improvements	• Considering technical requirements (devices, connectivity, integration) [10] • Considering EHR configuration [10] • Keeping the system regularly updated [18] • Developing standardized operating procedures and checklists, such as reminders to silence phone notifications and when conducting physical examinations [16] • Enhancing model recognition capabilities [17] • Developing a user-centred design approach and a system-thinking approach to overcome technical challenges [17] • Improving usability [16] • Improving documentation efficiency [16] • Ensuring AI tools are intuitive and capable of handling the dynamic nature of clinical environments [17]
Training Support	• Training and onboarding, ongoing support [3, 10, 16-18] • Providing comprehensive training [12, 16] • Promoting individualized training in using digital scribes [16]
Conducting Evaluation	• Establishing an iterative feedback mechanism for continuous improvement [12] • Ongoing evaluation of the technology's effectiveness and ensuring interoperability with existing EHR systems [10, 12] • Repeated or continuous evaluation [3, 14] • Conducting a thorough needs assessment through clinician interviews [12]
Organizational Preparation	• Considering institutional priorities, executive sponsorship, and internal champions to facilitate adoption [12] • Enhancing clinical workflows by emphasizing essential features [13, 16, 17] • Allowing the customization of verbal cues, and integrating seamlessly with existing EHR systems [18] • Developing policies to encourage the exploration and adoption of AI tools like digital scribes in clinical settings [17] • Developing operationalization and implementation plans with accountable, fair, and inclusive AI approaches to ensure the trustworthiness of the digital scribes [17] • Providing incentives for technology adoption [17]
Ethical Consideration	• Ensuring accuracy in transcribing and summarizing clinical encounters [18] • Maintaining data privacy and protection [18] • Upholding ethical standards [18] • Providing standardized consent processes with explanations on what the recordings will be used for, to help improve patients' comfort with recordings [16, 18] • development of evidence-based guidelines to ensure ethical and secure use of AI in healthcare [17]

8.3 Clinician outcomes

8.3.1 Reduced documentation burden

Studies suggest that using AI scribes can lead to reductions in documentation time both during and after clinical shifts[3, 12]. For example, a pilot study found that clinicians using virtual scribes reported an average time saving of 3 hours per shift[12]. Similarly, a simulated patient encounter study showed that using a digital scribe was 2.7 times faster than typing or dictating for the history section of a patient note, and 2.17 and 3.12 times faster for the physical exam section compared to typing and dictating, respectively[18]. In Haberle et al. study, the study group's 24-hour documentation deficiency rate dropped from 8.6% at baseline to 6.3% by the end of the study. Another study found that medical students summarizing mock consultations took less time when using a digital scribe system to generate an initial draft for editing than when summarizing manually[3].

8.3.2 Cognitive load

The introduction of AI scribes significantly reduced cognitive load across multiple dimensions, as evidenced by marked improvements in self-reported metrics. Mental demand decreased from 3.1 without the AI scribe to 1.8 with its use, indicating a substantial alleviation of cognitive strain. Similarly, physical demand dropped from 2.9 to 1.5, reflecting reduced physical effort in documentation tasks. Temporal demand also saw a notable reduction, from 3.4 to 1.8, suggesting better management of time pressures. Performance ratings improved significantly with the AI scribe, increasing from 2.9 to 4.1, highlighting enhanced efficiency and task outcomes. The effort required to complete tasks decreased from 3.8 to 2.2, and frustration levels fell from 4.1 to 2.3, indicating a more seamless and less stressful user experience. These results collectively demonstrate that AI scribes not only streamline documentation processes but also substantially lighten the cognitive burden on users. [11].

8.3.3 Improved provider well-being

While some studies haven't found statistically significant improvements in clinician burnout with AI scribe use, there is some evidence to suggest positive trends in provider engagement and reduced stress[10-12, 16]. Clinicians in one study reported feeling less stressed and more satisfied with their documentation time after using virtual scribes[12]. However, another study evaluating an AI scribe in a simulated emergency department environment, while finding improvements in cognitive and physical workload and provider wellness, acknowledged that the simulation may not have fully captured the complexity of real clinical practice[11].

8.3.4 Mixed impact on productivity

Studies examining AI scribes' effect on productivity have yielded mixed results. One study reported no significant changes in productivity measures like work relative value units (wRVUs) or panel sizes but acknowledged that external factors may have limited potential increases[10].

8.4 Implementation outcomes

8.4.1 Implementation and training needs

Successful integration of AI scribes into clinical workflows requires careful planning, implementation, and training, as seen in themes reported in **Table 5**.

Pilot testing and real-world evaluations were recommended to improve applicability and measure effects on workflow, care quality, and patient experience. Technical improvements should focus on robust infrastructure, seamless EHR integration, standardized procedures, and usability enhancements to ensure AI systems can adapt to dynamic clinical environments[12, 16]. A training support framework, emphasizing comprehensive and personalized onboarding, is critical for user adoption and proficiency [3, 10, 16-18]. Continuous evaluation was important and required iterative feedback loops, ongoing effectiveness assessments, and needs-driven refinements[3, 10, 12, 14]. Organizational preparation involves aligning AI adoption with institutional priorities, fostering executive sponsorship, and creating operational plans that emphasize accountability and inclusivity [13, 16, 17].

These themes provide a roadmap from the scientific literature for integrating digital scribes into healthcare with maximal impact and minimal disruption.

8.4.2 Patient perspectives and ethical considerations

Patient comfort and privacy are important considerations when implementing AI scribe technology[3, 13]. While some studies suggest that patients are generally accepting of AI scribes, authors are concluding that further research is needed to understand their perceptions and concerns [11, 13, 18]. Ethical considerations regarding data privacy and security, potential bias in AI algorithms, and the impact on the patient-physician relationship also need to be addressed [11, 13, 18].

8.5 Accuracy

8.5.1 Overall accuracy of AI Scribes

Accuracy is a key concern with AI scribes, as errors with significant clinical consequences were reported in studies[15, 17]. Some studies have identified limitations in the accuracy of AI-generated notes, particularly with non-objective data[10, 12, 14, 17]. Studies have also highlighted variability in the quality of notes produced by different scribes, whether human or AI[12]. The quality and consistency of AI-generated notes can be impacted by factors such as transcript quality and the complexity of the medical case[10, 17].

8.5.2 Characteristics influencing accuracy:

Systems with a human review step did not have a quantifiable effect on patient safety, patient experience, or clinical documentation [10]. In a simulated emergency department environment, participants rated the accuracy of the AI scribe-generated notes as 3.5 out of 5, on average, across all encounters [11]. A

system for generating medical scribes and prescriptions and found that the system-generated scribes were 87.5% similar to manually generated [13]. Similarly, the system-generated prescriptions were 96.2% similar to manually generated prescriptions[13].In a study using ChatGPT-4 on transcripts of simulated patient-provider encounters. Their study found that ChatGPT-4 had substantial variability in errors, accuracy, and note quality. The AI model's accuracy was lower for longer transcripts and more complex medical cases. The quality and reliability of the AI-generated notes were deemed not to meet the standards required for clinical use [14] Similarly, a multi-agent system using fine-tuned medical LLMs outperformed OpenAI's GPT-4o mini in generating summaries from patient conversation transcripts [15].

8.6 Results visualization

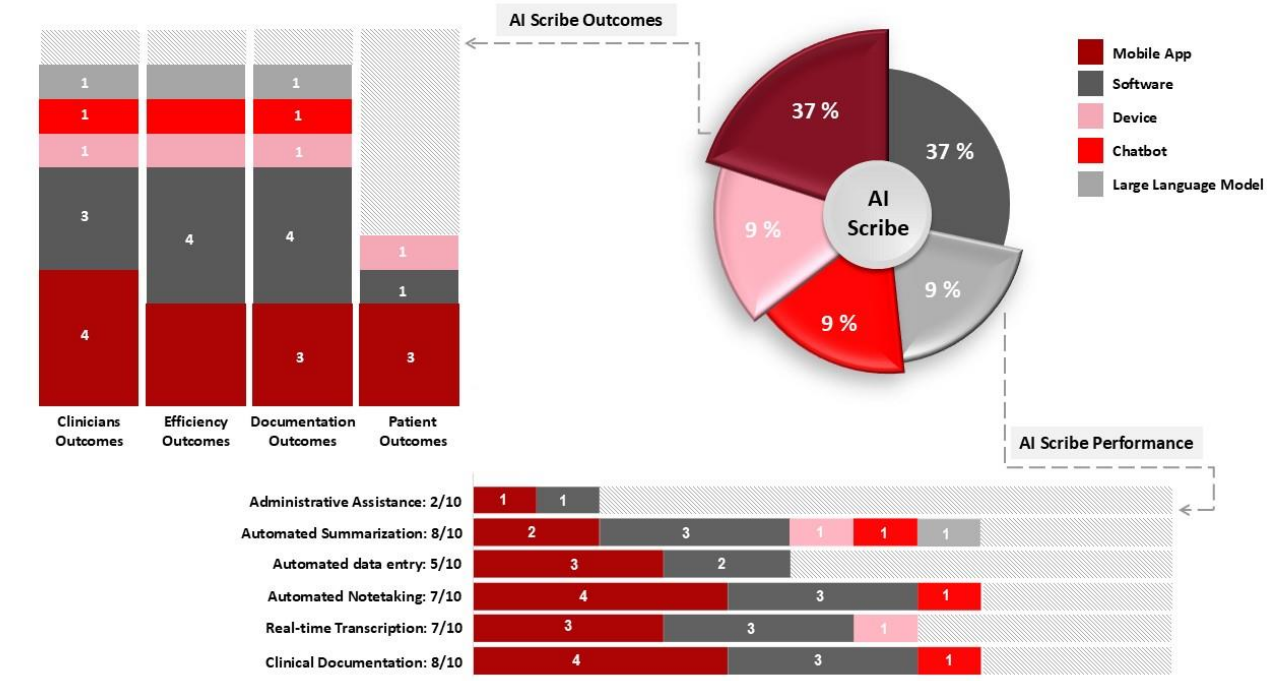

Figure 3. AI scribe systems, performance, and outcomes

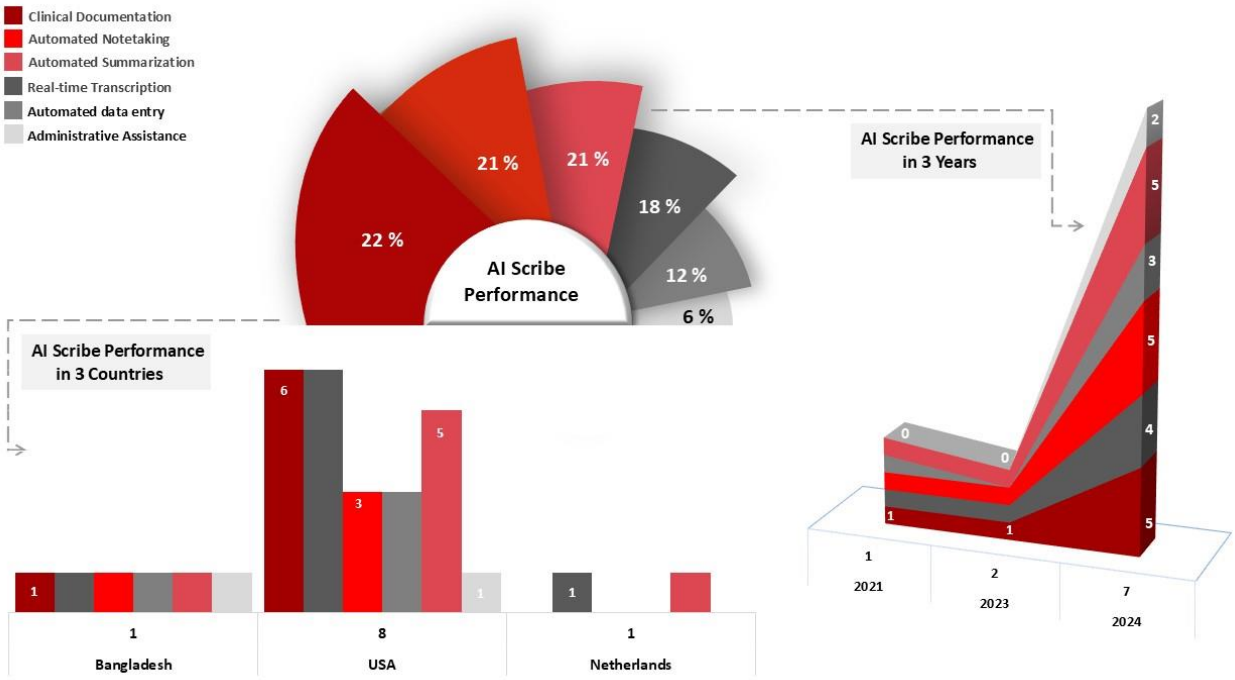

Figure 2. AI scribe performance in years and countries

9 Discussion and conclusions

9.1.1 Summary of results

The implementation of digital scribes in healthcare settings has shown mixed impacts on various aspects of clinical practice. While there are some positive trends observed in provider engagement and reduced workload, one source showed no improvements on physician burnout and unease from patients. Overall, while digital scribes offer potential benefits in specific areas like provider experience, further research is needed to determine their impact on productivity, documentation quality, patient experience and cost effectiveness.

Implementation of digital scribe technology, like any complex health technology system, requires careful planning and execution. A key aspect described by the studies was preparing the EHR system for integration with the digital scribe solution. This involved creating new note types, templates, and workflows specifically designed for notes generated by the digital scribe. Effective training for clinicians, engagement and commitment to training were identified as facilitators for adoption. Individualized training, particularly one-on-one champion training in the early stages, as shown to improve the learning curve for providers. Ongoing support, such as an on-site support person, is also important for troubleshooting any technical or workflow issues that arise during the implementation. Addressing these aspects thoroughly during implementation can enhance the effectiveness of digital scribe solutions in clinical practice.

The studies suggest that AI Scribes accuracy can vary depending on the specific system and context. Some systems, especially those that incorporate human review, may achieve high levels of accuracy, while others still require further development to improve reliability.

9.1.2 Strengths of this study and Impacts of the results

This study is the first knowledge synthesis on the impacts of AI scribes. Our sources from 2023 and 2024, and the inclusion of preprint sources offers an up-to-date and comprehensive portrait of the rapidly evolving landscape of AI Scribe interventions. We thus identified both current trends and emerging developments. We also identified interventions leveraging cutting-edge language models from 2024, such as GPT-4o, underscoring our timely contribution to advancing our understanding of the impacts of AI scribes.

9.1.3 Limitations

As the accuracy of AI scribes can vary, it is crucial to incorporate human review, especially in complex medical cases, to ensure patient safety and reliable documentation. Many of the studies involve small sample sizes and specific healthcare settings, which may limit the generalizability of their findings to other contexts. The accuracy and performance of AI scribe systems can vary greatly depending on the specific technology, model training data, and implementation approach. Many studies primarily focus on proximal outcomes, such as documentation time and perceived efficiency gains. Longer-term outcome assessments are needed to demonstrate the sustained impact of AI scribes on provider well-being, patient care quality, and overall healthcare costs. Most sources concentrate on the physician experience and do not explore patient perspectives on the use of AI scribes. It is important to also consider patient privacy concerns, potential biases in AI algorithms, and the impact of AI scribes on the patient-physician relationship. The sources focus mainly on the technical aspects and performance of AI scribes, with limited focus on the potential ethical and social implications. Further research is needed to address concerns related to data privacy, algorithmic bias, job displacement, and the changing dynamics of the healthcare workforce in the context of AI adoption.

9.1.4 Conclusions

Physicians are increasingly burdened by excessive documentation demands, diverting valuable clinical time and contributing to burnout and decreased job satisfaction. AI scribes offer a promising solution by alleviating these tasks, enabling physicians to focus more on patient interactions. This shift not only increases face-to-face time but also enhances the quality of patient care and improves the overall patient experience. Our review highlights a consistent trend of positive impacts on clinicians, alongside actionable insights to optimize the implementation of AI scribes within clinical workflows and existing technological infrastructures.

As the adoption of thoughtfully integrated AI scribes continues to grow, future research will provide valuable data on long-term outcomes, deepening our understanding of their effects on patient care, cost-effectiveness, and environmental sustainability. These findings will guide the development of more efficient and impactful AI-driven solutions in healthcare.

10 References

1. Johnson, K.B., M.J. Neuss, and D.E. Detmer, *Electronic health records and clinician burnout: a story of three eras.* Journal of the American Medical Informatics Association, 2021. **28**(5): p. 967-973.
2. Gardner, R.L., et al., *Physician stress and burnout: the impact of health information technology.* Journal of the American Medical Informatics Association, 2019. **26**(2): p. 106-114.
3. van Buchem, M.M., et al., *Impact of a Digital Scribe System on Clinical Documentation Time and Quality: Usability Study.* JMIR AI, 2024. **3**: p. e60020.
4. Tierney, A.A., et al., *Ambient artificial intelligence scribes to alleviate the burden of clinical documentation.* NEJM Catalyst Innovations in Care Delivery, 2024. **5**(3): p. CAT. 23.0404.
5. Quiroz, J.C., et al., *Challenges of developing a digital scribe to reduce clinical documentation burden.* NPJ digital medicine, 2019. **2**(1): p. 114.
6. Higgins, J.P., et al., *Cochrane handbook for systematic reviews of interventions*. 2019: John Wiley & Sons.
7. Garritty, C., et al., *Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews.* Journal of Clinical Epidemiology, 2020.
8. Hong, Q.N., et al., *Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study.* Journal of clinical epidemiology, 2019. **111**: p. 49-59. e1.
9. Covidence, *Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia.*
10. 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.
11. Hata, R., et al., *117 Evaluating the Accuracy and Provider Wellness Impact of an Ambient Artificial Intelligence Scribe in a Complex Simulated Emergency Department Environment.* Annals of Emergency Medicine, 2024. **84**(4): p. S54-S55.
12. Hudelson, C., et al., *Selection and Implementation of Virtual Scribe Solutions to Reduce Documentation Burden: A Mixed Methods Pilot.* AMIA Summits on Translational Science Proceedings, 2024. **2024**: p. 230.
13. Islam, M.N., S.T. Mim, T. Tasfia, and M.M. Hossain, *Enhancing patient treatment through automation: The development of an efficient scribe and prescribe system.* Informatics in Medicine Unlocked, 2024. **45**.
14. Kernberg, A., J.A. Gold, and V. Mohan, *Using ChatGPT-4 to Create Structured Medical Notes From Audio Recordings of Physician-Patient Encounters: Comparative Study.* J Med Internet Res, 2024. **26**: p. e54419.
15. Lee, C., S. Kumar, K.A. Vogt, and S. Meraj, *Improving Clinical Documentation with AI: A Comparative Study of Sporo AI Scribe and GPT-4o mini.* arXiv preprint arXiv:2410.15528, 2024.
16. Nguyen, O.T., 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, 2023. **7**: p. e2200166.
17. Sezgin, E., J.W. Sirrianni, and K. Kranz, *Evaluation of a Digital Scribe: Conversation Summarization for Emergency Department Consultation Calls.* Appl Clin Inform, 2024. **15**(3): p. 600-11.
18. Wang, J., et al., *A patient-centered digital scribe for automatic medical documentation.* JAMIA Open, 2021. **4**(1): p. ooab003.

11 APPENDIX 1 - Strategies for identifying data sources

Medline (Ovid)

Date of the search: 01-11-2024

Database limit: no database search limit has been applied.

#	Search strategy	Results
1	((Ambient OR Clinical OR Medical OR automated OR virtual OR Digital OR AI OR "Artificial Intelligence" OR E) adj1 scrib*).ti,ab,kf	189

Embase (Embase.com)

Date of the search: 01-11-2024

Database limit: results have been limited to Embase & prepublications only.

#	Search strategy	Results
1	((Ambient OR Clinical OR Medical OR automated OR virtual OR Digital OR AI OR "Artificial Intelligence" OR E) NEAR/1 scrib*):ti,ab,kw	254
2	#1 AND [embase]/lim NOT ([embase]/lim AND [medline]/lim)	97

CINAHL (EBSCO)

Date of the search: 01-11-2024

Database limit: no database search limit has been applied.

#	Search strategy	Results
1	TI ((Ambient OR Clinical OR Medical OR automated OR virtual OR Digital OR AI OR "Artificial Intelligence" OR E) N1 scrib*) OR AB ((Ambient OR Clinical OR Medical OR automated OR virtual OR Digital OR AI OR "Artificial Intelligence" OR E) N1 scrib*)	144

Web of Science

Date of the search: 01-11-2024

Database limit: no database search limit has been applied.

#	Search strategy	Results
1	TS=((Ambient OR Clinical OR Medical OR automated OR virtual OR Digital OR AI OR "Artificial Intelligence" OR E) NEAR/1 scrib*)	321
2	TS=(health* OR care OR Medic* OR clinical)	11,720,971
3	#1 AND #2	233

Arxiv (https://arxiv.org/)

Date of the search: 01-11-2024

Database limit: search in all fields.

#	Search strategy	Results
1	"Ambient scribe" OR "Clinical scribe" OR "Medical scribe" OR "automated scribe" OR "virtual scribe" OR "Digital scribe" OR "AI scribe" OR "E scribe"	5

Google Scholar (https://harzing.com/resources/publish-or-perish)

Date of the search: 01-11-2024

Database limit: only up to the 100 first results have been considered; citations and patents options have been removed.

#	Search strategy	Results

1	"Ambient scribe" OR "Clinical scribe" OR "Medical scribe" OR "automated scribe" OR "virtual scribe" OR "Digital scribe" OR "AI scribe" OR "E scribe"	100