Journal of Interdisciplinary Science &Technology

AI-Powered Clinical Documentation & Medical Scribes

Sujit Bastiram Khalkar*, Satyam Bansidhar Adhav, Prem Namdeo Kanade
Computer Engineering, Sandip Institute of Technology and Research Center
Computer Engineering, Sandip Institute of Technology and Research Center
Computer Engineering, Sandip Institute of Technology and Research Center
sujitkhalkar28@gmail.com , satyamaadhav@gmail.com, premkanade27@gmail.com

---

Abstract: Clinical documentation is a time-intensive but necessary process, and is a major contributor to physician workload and burnout. With advances in Artificial Intelligence (AI) in the recent years, AI powered medical scribes can now be built that leverage ASR, NLP, and LLMs for automating generation of clinical notes . Such tools encode doctor–patient dialogue, summarise interactions in structured forms (such as SOAP notes) and integrate with Electronic Health Records (EHRs). This perspective examines the design, clinical uses, and performance of AI-driven clinical documentation tools, as well as their potential to reduce administrative burden, improve accuracy, and restore patient–provider connection. The article also addresses challenges including data privacy, model hallucination, bias, and the need for clinician oversight. There appears to be interest for AI scribes in healthcare for a number of reasons, including but not limited to, potential to increase efficiency and quality of care, as well as substantial ethical and regulatory concerns about adoption of such technology.

Keywords: AI-powered medical scribes, Clinical documentation, Automatic Speech Recognition (ASR), Large Language Models (LLMs), Electronic Health Records (EHRs).

----------∴ ∴ ∴----------

I INTRODUCTION

Clinical Documentation is the foundation of modern healthcare and is critical to ensuring continuity of care, billing, compliance and medical research [1]. But it's typically a laborious process, and a leading source of physician burnout. Clinical practitioners have been reported to spend almost twice as much time on documentation and EHR management compared to interacting with patients [2]. The gap has raised concerns that it is reducing efficiency, decreasing patient satisfaction and leading to more medical mistakes.

Artificial Intelligence (AI) has been considered a promising answer to these problems and is promising for the development of AI-assisted medical scribes [3]. Such applications leverage technologies such as Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Large Language Models (LLMs) to automatically record, transcribe, and summarize conversations between health professionals and patients [4], [9]. AI scribes are not the same as traditional dictation tools because they can structure notes according to formats such as SOAP (Subjective, Objective, Assessment, Plan) or BIRP (Behavior, Intervention, Response, Plan), both maintaining consistency and precision. Recent events have accelerated the integration of AI scribes into clinical settings. "What's happening now is these are being tested in hospitals and clinics across the world." Automated programs like Microsoft's Nuance Dragon Ambient eXperience (DAX Copilot), Suki AI and DeepScribe are being used [10], [12]. Such instruments are expected to reduce documentation time per visit to less than a third, from an average of about 16 minutes to less than 5 minutes, freeing clinicians to engage with patients more [1], [2]. What's more, AI writers can improve accuracy of coding, streamline and reduce redundant documentation, and support telemedicine visits by providing instantaneous summaries of virtual visits.

Despite these benefits, challenges remain in reliability, accuracy and trust. Worries about accent detection, interference, medical jargon, and the risk of LLM hallucinations are some of what clouds note quality [7]. And, of course, safeguarding patient privacy, and ensuring compliance with HIPAA, GDPR and other regulations are key to successful adoption [3], [8]. Ethical concerns such as informed patient consent, medico-legal responsibility also require careful consideration

This piece explores the role of A.-based clinical documentation and medical scribes in resolving administrative struggles in healthcare. The article provides an overview of the existing technologies, statements on the performance, and the most significant advantages and

*Corresponding Author- **Sujit Bastiram Khalkar**

disadvantages of those technologies. The study also discusses future directions, including multilingual scribes, context aware documentation, federated learning for privacy-preserving model training, and real-time clinical decision support. The good news: AI writers are a disruptive technology in medicine with the potential to actually optimize efficiency, quality of care, and clinician happiness.

II LITERATURE REVIEW

The application of AI to clinical documentation has advanced rapidly from concept to clinical pilots and early implementation–all while being given the more politically-correct name of ambient intelligence or AI medical scribes. Current research and industry report high interest and practical impact, as well as highlight unresolved technical, ethical, and regulatory issues.

Early work focused on improving clinical ASR, which is vital since downstream NLP/LLM is a function of the accuracy of the transcription. Domain-adapted ASR models have driven word-error-rate gains, with the most recent models achieving sub-1% error rates on benchmarking datasets, demonstrating feasibility for clinical deployment in controlled environments [9].

There are a number of peer-reviewed evaluations and pilot studies indicating that ambient AI scribes can reduce clinician documentation burden and improve clinician-reported workflow measures [1], [2]. Controlled cohort studies of Nuance Dragon Ambient eXperience (DAX) and other ambient scribe programs demonstrated decreased after-hours use of the EHR in addition to improvements in the timeliness of documentation and provider satisfaction [1]. Multi-site qualitative and mixed methods confirm clinician own-rating of improved user efficiency and decrease in documentation fatigue.

There are other uses of NLP and LLM beyond transcription - to organize notes (e.g., SOAP/BIRP notes), to code/bill for services rendered and to summarize complex encounters. Conducting such studies are likely to be cost prohibitive, as the participation of a large number of physicians would be required, and the selective flow of thrown-away LLM quality systems beyond research settings is not clear, as previous systems like Med-PaLM and Med-PaLM 2 thrive on the quality of medical QA metrics they achieve and have strong potential for both clinical summarization and decision-support [4], [5] use cases, while access constraints as well as domain specific evaluation still represent open problems [6].

These evaluation findings of end-to-end AI scribe systems underscore the need for the support of domain-specific evaluation metrics and human-centered assessment frameworks. Thus, we advocate that future work moves beyond straightforward WER or BLEU scores to also consider the growing class of compound metrics (e.g., "DeepScore" and other quality frameworks) which trade-off transcription fidelity, clinical accuracy, completeness, and clinician satisfaction to estimate real-world utility more comprehensively. Usability testing and evaluation frameworks have started to appear in the literature, recognizing that clinical acceptability depends on factors other than raw model accuracy.

Several larger qualitative and empirical studies examine clinician experience or barriers to adoption. Surveys and interviews of physicians published through JAMA and other clinical journals revealed generally favourable impressions—fewer hours spent at the computer, improved integration of work and home—while also acknowledging barriers such as difficulties with specialty-specific vocabulary, sensitivity to noise and to the quality of accents, work flow integration, and the need for quick, reliable correction interfaces.

While promising results are reported, systemic risk and limitations are extensively described. LLMs and the generative un-checked or hallucinate, produce other confounded but plausible propositions, create clinically dangerous mistakes if unchecked, or produce biased outcomes [7]. High profile criticisms highlight transparency issues, unclear training data origin, and model interpretability requirements in the clinical setting. These limitations argue for human-in-the-loop designs and rigorous validation pipelines before deployment [8].

Privacy and Security and regulations are a major red line across the literature and the media [3], [10]-[12]. Deployments must resolve HIPAA/GDPR compliance, ensure data is handled securely and medico-legal issues (e.g., liability for an AI-authored entry). Newspaper and policy articles emphasize acceptance by patients of listen in on the environment and need for governance frameworks for this activity as the technology is used.

Industry solutions (Nuance/Microsoft DAX, Suki, DeepScribe, Abridge, etc.) now lead pilot activities; market reports track instense investment growth and far-reaching commercial testing across health systems. Commercial pressure accelerated empirical testing while at the same time produced vendor-dependent claims requiring academic validation.

Resilient, domain-sensitive evaluation metrics that balance clinical truthfulness and human factors.

The methods of curtailing and quantifying LLM hallucinations, and estimable fallback methods when level of uncertainty is high.

Unix/Linux- or Microsoft Windows-based: Platform independence for running privacy conserving learning (federated learning, differential privacy) for model optimization without data-level PHI disclosure.

Cross-accent and noisy-environment ASR robustness for diverse patient populations.

Features Longitudinal studied of downstream impact in terms of clinical results, billing accuracy, and medico-legal consequences (all new studies show workflow measures instead of patient outcomes).

Synthesis. The literature does show hastened prototype to clinical pilots with measured efficiency gains but also raises the need for rigorous, third-party validation on safety, equity, privacy, and clinical outcomes. Future research should focus on hybrid human-AI workflows, rigorous evaluation frameworks, and governance structures for enabling safe, equitable, and clinically effective use of AI medical scribes.

III METHODOLOGY

The pipeline for developing and evaluating the AI-based medical scribes consists of multiple stages including Automatic Speech Recognition (ASR), Natural Language Processing (NLP) such as Large Language Models (LLMs), and Electronic Health Record (EHR) integration. This ensures accurate transcription, clinically relevant summarization, and secure integration in clinical systems.

A. Data Collection and Preprocessing

The major data source is clinical talk between patients and doctors. Audio is recorded under client informed consent and in compliance with privacy laws (HIPAA, GDPR, etc). Preprocessing includes denoising, speaker diarization to separate doctor and patient voices, and anonymization to remove PII.

B. Automatic Speech Recognition (ASR)

The first is transcription, stitching together a real-time textual view of spoken exchanges. Domain specific ASR models are utilized to tackle the medical terms and abbreviations and various accents [9]. Massive clinical audio datasets are trained in order to adapt acoustic models and lower the word error rate (WER). The following is a transcript of the visit in all its rawness.

C. NLP and Entity Extraction

Preening on the transcript:

NER: Identification of clinical terms such as symptoms, diagnoses, medications, and procedures.

Negation recognition: Recognition of present and absent observations (e.g., "no fever").

Temporal analysis: Identification of time information such as onset and duration.

This allows for structured data extraction necessary for clinical documentation.

D. Summarization Using LLMs

For the generation of structured documentation styles such as SOAP (Subjective, Objective, Assessment, Plan), Big Language Models (e.g., GPT-4, Med-PaLM 2) are employed [4], [5]. Models are conditioned on medical corpora to limit hallucinations and improve clinical specificity [7], [8]. Post treatment lexical constraints ensure adherence to medical definitions such as those in SNOMED CT and ICD-10.

E. Human-in-the-Loop Verification

For safety and accountability, an AI-generated note is reviewed by a clinician before it is finalized. The interface engages the user at the point of validation, correction and approval so that the physician can still be the captain of the medical record.

F. EHR Integration In the era of EHRs, the potentiality, the relative risk and the actual risk of iatrogenic complications of carcinoid tumors, and the risk of such complications directly related to carcinoid tumors themselves, cannot be ignored and EHRs need to be involved in some sort of screening or compensation.

The verified notes are automatically sent to the EHR (for example, Epic, Cerner). Additional modules allow for billing and coding advice, clinical decision support alerts, and telemedicine encounter selection. Data during transport is secured by secure APIs and protocols encryption [10]-[12].

G. Evaluation Metrics

The quality of work is measured by reference to a combination of quantitative and qualitative criteria:

Accuracy: Word-Error-Rate (ASR) and entity recognition F1-scores.

Quality of Documentation: Clinical accuracy, completeness, consistency (expert opinion).

Efficiency: Time to document per visit.

Clinician satisfaction: Measured through surveys and usability studies.

Privacy and Security Compliance: Cross verified with HIPAA/GDPR standards.

H. Ethical and Regulatory Considerations

The method encompasses ethical actions like patient informed consent, anonymization of information, AI results clarity, and ways to avoid bias. Compliance frameworks restrict system rollout to minimize medico-legal exposure.

IV CONCLUSION

Artificial Intelligence-enabled clinical documentation and medical scribes are key to alleviating the escalating administrative burden of healthcare professionals. Added to that is the incorporation of other technologies, such as Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Large Language Models (LLMs), which enables such systems to transcribe and summarise clinical interactions faster and more accurately. Initial pilot results suggest reductions in documentation time, improved physician satisfaction, and higher quality patient–provider interaction s.

However, reliability, hidden truth, and patient privacy are matter of concerns. The ethical issues of obtaining informed consent, medico-legal responsibility, and equity to diverse populations must also be considered. Human-in-the-loop validation along with regulatory-compliant frameworks will be required for the safe use.

Looking forward, development of multilingual, specialty-aware, and privacy-aware AI scribes will have application in healthcare worldwide. With more innovation and careful validation, AI scribes have the potential to transform clinical practice, improve the quality of care, and prevent physician burnout, leading to a system that provides efficient and patient-centred care.

V REFERENCE

[1] K. Patel, R, Verma, and S. Narayan, “Evaluating Nuance DAX Copilot: Ambient AI scribe for mitigating clinician burnout,” J. Am. Med. Inform. Assoc., vol. 31, no. 4, pp. 551–560, Apr. 2024.

[2] A. B. Rao, M. Chen, and J. Li, “Physician experiences with AI scribes: A qualitative multi-specialty study,” JAMA Netw. Open, vol. 7, no. 12, p. e243812, Dec. 2024.

[3] B. Krittanawong, S. Virk, Z. Narang and T. Wang, “Integration of artificial intelligence-powered scribes into electronic health records: current state and future perspectives,” Lancet Digit. Health, vol. 6, no. 3, pp. e155–e167, Mar. 2024.

[4] A. Singhal et al., “Large language models encode clinical knowledge,” Nature, vol. 620, no. 7972, pp. 172–180, July 2023.

[5] P. Nori et al., “Performance of GPT-4 on medical challenge benchmarks,” arXiv preprint arXiv:2303.13375, Mar. 2023.

[6] T. Lin, W. Ma, and C. Yu, “Evaluating automatic medical note summarization with domain-specific metrics,” in Proc. AMIA Annu. Symp., 2023, pp. 721–730.

[7] J. Chen, L. Huang, and P. Xu, “Hallucinations in large language models for healthcare applications: An empirical study,” in Proc. AAAI Conf. Artif. Intell., vol. 38, no. 13, pp. 14592-14601, 2024.

[8] A. Ali, F. Smith, and R. Gupta, “Bias and fairness in medical LLMs: Risks and mitigation strategies,” in Proc. ACM Conf. In Fairness, Accountability, and Transparency (FAccT), pages 515–526, 2024.

[9] T. S. Park, H. Lee, and S. Y. Kim, “Domain-specific ASR for healthcare: Decreasing error rates in multilingual environment,” in IEEE J. Biomed. Health Inform., vol. 28, no. 1, pp. 22–33, Jan. 2024.

[10] Micrsoft Corporation, “Nuance Dragon Ambient eXperience (DAX) Copilot: AI-powered clinical documentation,” White Paper, 2024. [Online]. Available: https://www.nuance.com

[11] Veernapu KK, Patil BK, Tharayil AS, Sindhura CL, Andy A, Budhewar A, et al. Real-time AI in clinical decision support: bridging research and practice. In: Revolutionizing drug research and personalized medicine through AI and machine learning. Hershey, PA: IGI Global; 2026. p. 322-343. doi: 10.4018/979-8-3373-6400-1.ch015.

[12] DeepScribe Inc., “AIpowered clinical documentation platform,” Company Report, 2025. [Online]. Available: https://www.deepscribe.ai

