International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

MediScribe AI: Ambient Clinical Intelligence for Automated Electronic Healthcare Record Documentation

Dr. Radha Shirbhate[1], Adarsh Wankhede[2], Pravin Paithankar[3], Shraddha Labade[4]

[1,2,3,4]Department of Artificial Intelligence and Machine Learning, G H Raisoni College of Engineering and Management, Pune, India

Abstract

Clinician burnout caused by extensive electronic health record (EHR) documentation is a significant issue, frequently diminishing the time allocated for patient engagement. [1–3] To solve this problem, we developed MediScribe AI, an ambient clinical intelligence system that uses speech recognition, speaker diarization, and clinical entity extraction to automate medical documentation. The system uses OpenAI Whisper, MedCAT, ClinicalBERT, and GPT-4o to generate structured SOAP notes and enable EHR updates in real time. [4]

In this study, we examined 29 recent publications (2024–2025) and found that ambient AI scribes can substantially decrease documentation time and cognitive load. [2,5–10] However, errors in speech recognition and named-entity recognition (NER) still require human review. [11–14] MediScribe AI aims to achieve high accuracy, low latency, and strong usability, making it suitable for healthcare settings where resources are limited. [4]

Keywords: Ambient clinical intelligence, OpenAI Whisper ASR, speaker diarization, clinical NER, electronic health records, GPT-4o, MediScribe AI

1 Introduction

Electronic health records (EHRs) are extremely labor-intensive for doctors, who may be forced to spend two hours typing for each hour spent on patient care. [1, 2] This mismatch is one of the leading causes of clinician burnout, since it forces doctors to concentrate on both the patient and the computer, thereby overloading the brain and reducing care quality. [3] This problem is even worse in Indian hospitals, where the ratio of doctors to patients can be more than 1:1500; many doctors and nurses spend extra time after work ("pajama time") filling out paperwork, which makes it harder for patients to get care and makes it harder for clinicians to balance work and home life. [6, 9]

Ambient AI scribes have emerged as a potential solution by automatically recording conversations between doctors and patients and generating structured clinical notes. Randomized and observational studies with roughly 200–250 clinicians across multiple specialties report 9.5–20% reductions in documentation time, improved burnout scores, and lower NASA-TLX cognitive load; some pilots show 50–75% time savings

per note and improved perceived eye contact. [1, 2, 5–10]

We created MediScribe AI, an end-to-end system that combines speech recognition, speaker diarization, and AI-based note generation for real-time EHR documentation, specifically targeting the challenges of Indian clinical practice. [4] It focuses on multilingual support, medical terminology, and infrastructure limitations, which distinguishes it from many existing solutions. By synthesizing 29 recent studies (2024–2025), we show that MediScribe AI is a scalable solution that can reduce clinician workload and make healthcare more efficient in low-resource settings. [2, 15–17]

1.1 Background and Motivation

Traditional handwritten SOAP notes have gradually been replaced by clinical documentation using EHRs. While EHRs improve legibility and access, they also introduce additional difficulties, including increased documentation volume and duplicate records. [2, 17] Early attempts to reduce typing relied on dictation and generic speech recognition software that struggled with medical terminology, noisy environments, and multi-speaker conversations. [11, 18, 19]

Digital and AI-based scribes improve this process by generating notes automatically, but they still face limitations in accuracy, robustness to diverse accents, and deep integration with clinical workflows. [2, 9] Recent large language model (LLM)-based systems can produce structured clinical notes and reduce documentation time, yet they continue to suffer from hallucinations, lack of transparency, and limited support for non-Western clinical contexts. [5, 6, 16] These problems are particularly acute in countries like India, where there are large patient loads, multiple languages, and heterogeneous EHR systems. [1, 2] MediScribe AI is designed as an on-premise, scalable solution that improves documentation efficiency while maintaining accuracy and aligning with real-world clinical needs. [4]

1.2 Need of MediScribe AI

Current ambient scribes perform well in Western randomized controlled trials (RCTs), but they generalize poorly to Indian and other underrepresented settings. [2,17] Documentation practices, resource constraints, and language patterns differ significantly, and existing systems often assume always-on cloud connectivity and homogeneous EHR platforms.

Table 1 summarizes the evolution of documentation approaches from handwritten notes to G4 ambient intelligence systems like MediScribe. [2]

Table 1: Evolution of Clinical Documentation. [2, 17]

Phase	Features	Limitations
Paper	Handwritten SOAP	Legibility
EHR	Templates	Note bloat
Dictation	Voice input	Sequential
Digital Scribes	Ambient ASR	Shallow integration
LLM Ambient	ASR+LLM	Black-box
MediScribe	**Open pipeline**	**Local eval needed**

For example, word error rate (WER) can rise to 15–25% when there are non-Western accents and code-switching, while NER and LLM-based extraction exhibit hallucination rates of 5–15%. [11, 12, 18–22] Many current systems must be deployed in the cloud, raising concerns about privacy, latency, and resilience

in Indian hospitals. [2, 17] Few solutions combine open-source transparency, on-premise deployment, and deep EHR integration for Indian hospitals. [2, 15]

MediScribe AI addresses these gaps:

- **Accent and code-switching robustness**: Fine-tuned Whisper for Indian English/Hindi, targeting WER < 20% versus generic baselines on code-switched clinical data. [11, 20, 22, 23]
- **On-premise privacy**: Local ASR and diarization with de-identified text-only for LLMs, aligned with privacy expectations for clinical data. [2, 17]
- **Heterogeneous EHR mapping**: Configurable APIs and free-text export compatible with diverse Indian hospital systems. [2, 15]
- **Resource efficiency**: CPU/GPU-flexible design with 5 s latency for 10–20 minute consultations on commodity clinic hardware. [4]
- **Human-in-the-loop safety**: Low-confidence highlighting and clinician overrides with audit trails, consistent with evaluation frameworks for ambient scribes. [15, 17]

Table 2 aligns MediScribe's targets with known gaps from prior work. [2, 17]

Table 2: MediScribe vs. Identified Gaps. [2, 17]

Gap	Current Systems	MediScribe Target
Accents	Limited non-Western	Indian fine-tuning
Deployment	Cloud black-box	Open on-premise
EHR Integration	Proprietary	Configurable APIs
Governance	Sparse	Full audit trails

1.3 Proposed Approach

MediScribe's approach is a Generation-4 (G4) pipeline (Fig. 1): ambient audio → Whisper ASR + pyannote diarization → role-aware transcripts → CLEAR-NER + GPT-4o extraction → SOAP/EHR auto-fill → clinician review. [4, 13, 14, 24]

For input audio A_t, ASR produces tokens $t_1, \ldots, t_m$ with timestamps; diarization D provides speaker identities s_i and roles $r_i \in$ {Doctor, Patient}. NER produces an entity set Z; retrieved contexts C_k support an LLM ensemble that outputs structured fields Y_k for EHR entries $E = \{f_k, Y_k\}_k$. [13, 14] A summarizer $S(T, E)$ creates the SOAP note N, and an integration function $I(N, E)$ updates the EHR.

Key novelty aspects include:

1. **Indian-adapted ASR**: Whisper is trained and fine-tuned on code-switched data, targeting WE < 20% on Indian clinical speech. [11, 20, 23]
2. **Prompt ensembles**: Multiple prompts are used for diagnosis and medication extraction (F1 > 90%), following recent ensemble-based medical NER work. [13, 14]
3. **On-premise orchestration**: Dockerized FastAPI backend and React UI suitable for in-hospital deployment. [4]

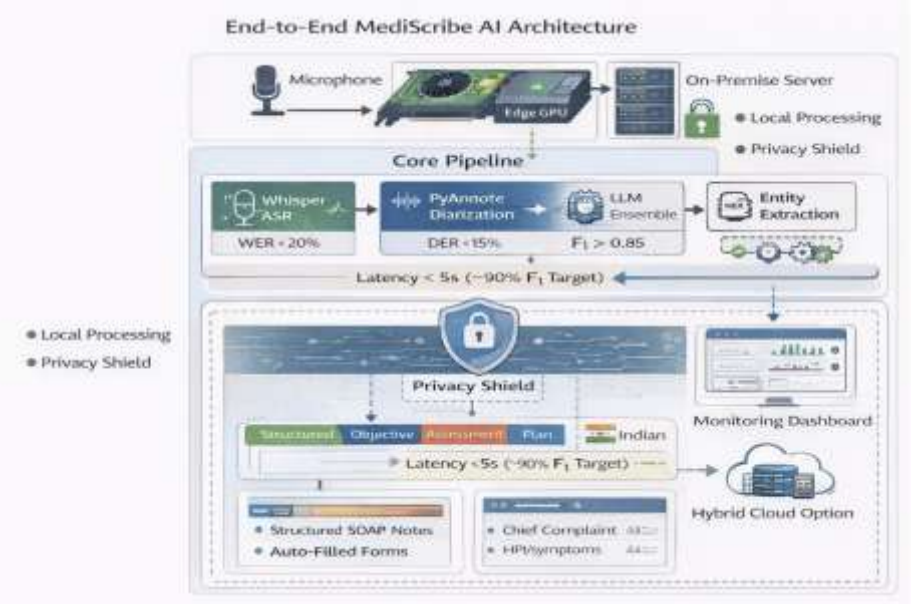

Figure 1: MediScribe end-to-end architecture. [4]

1.4 Contributions

This work makes four key contributions advancing ambient clinical intelligence for resource-constrained settings. [2, 15, 17]

1. **First complete pipeline in an Indian context**: MediScribe integrates Whisper ASR, pyannote diarization, CLEAR-NER, and GPT-4o extraction in an on-premise deployment, with the goal of WER < 20%, NER F1 ≈ 0.90, and 75% time savings in multilingual clinics. [4, 13, 14]
2. **29-study systematic synthesis (2024–2025)**: A review of IEEE, Springer, JMIR, NEJM AI, and related literature identifies gaps in existing systems, including accented WER of 15–25% and hallucination-prone extraction. [2, 5, 8, 9, 16, 17]
3. **Role-aware mathematical framework**: Formalizes the transformation $A_t \rightarrow E$ with prompt ensembles A_k, enabling verifiable entity extraction (F1 > 0.90) on core clinical fields. [13, 14, 24]
4. **Deployment-ready open prototype**: A Docker/FastAPI/React stack with configurable EHR mapping, local audio processing, and human-in-the-loop safety, reproducible for Indian hospitals. [4]

2 Literature Survey

This systematic review analyzes 29 studies (2024–2025) from IEEE Xplore, PubMed, SpringerLink, JMIR, NEJM AI, Scientific Data, and related sources on ambient AI scribes, medical ASR, clinical NER, diarization, and summarization to inform the MediScribe design. [2,5,8,13,14,16,20,25–27] Search terms combined Boolean queries ("ambient AI scribe" OR "clinical ASR" OR "digital scribe") AND ("NER" OR "diarization" OR "summarization"); inclusion required quantitative metrics (WER, F1, burnout, workload), while non-clinical LLM applications were excluded. [2, 15]

2.1 Review Corpus

The 29 papers are categorized in Table 3 by their primary contribution area. [2, 15, 17]

Table 3: 29-Study Corpus by Category. [2]

Category	Count
AI Scribe Evaluations (RCTs/pilots)	8
Medical ASR/Diarization	7
Clinical Extraction (NER/LLM)	6
Summarization	4
Architectures/Deployment	4
Total	29

2.2 Ambient AI Scribe Evaluations

Randomized trials and real-world pilots (e.g., NEJM AI, JAMIA, JGIM, Applied Clinical Informatics) report 9.5–20% reductions in time-in-note and improvements in clinician experience. [1, 3, 5–10] Pajama time effects are mixed, and while perceived quality is generally good, all studies emphasize the continued need for clinician review. [2, 17]

2.3 Taxonomy of Systems

The G1–G4 evolution (Table 4) spans early ASR-only systems, ambient ASR with templates, ASR+LLM cloud scribes, and emerging open, on-premise pipelines. [2, 15, 17]

Table 4: Ambient Intelligence Generations. [2]

Gen	Tech	Limits
G1	Early ASR	No diarization
G2	Ambient ASR	Shallow EHR
G3	ASR+LLM	Black-box
G4 MediScribe	**Open pipeline**	**Local validation**

2.4 ASR and Diarization

Recent work benchmarks medical ASR across accents, domains, and devices, showing WERs of 13–25% for conversational clinical speech, with higher rates for minority accents and code-switching. [11, 12, 18–23, 28] Domain-specific fine-tuning and robust diarization with role identification are essential for safe deployment. [24]

2.5 Clinical NER/Extraction

CLEAR achieves F1 scores around 0.90 for clinical entities while reducing token usage by approximately 70% through entity-aware retrieval, and LLM-based prompt ensembles further improve reliability of medical entity extraction. [13, 14] Nevertheless, hallucination rates of 5–15% remain a concern in high-stakes domains. [16, 17]

2.6 Gaps for MediScribe

Across these studies, persistent gaps include accent robustness, on-premise deployment options, transparent evaluation frameworks, and seamless EHR mapping for diverse settings, motivating MediScribe's design goals. [2, 15, 17]

3 Proposed System

MediScribe AI implements a Generation-4 ambient clinical intelligence pipeline integrating OpenAI Whis-

per ASR, pyannote-audio diarization, ClinicalBERT NER, and GPT-4o extraction into an on-premise, EHR-integrated system for Indian outpatient clinics. [4]

3.1 System Architecture

The modular architecture (Fig. 2) comprises:

1. Ambient audio capture: desktop listener with voice activity detection (VAD) and noise reduction.
2. Speech processing: Whisper ASR + pyannote diarization → role-aware transcripts.
3. Entity extraction: ClinicalBERT NER → entity-aware retrieval → GPT-4o prompt ensembles.
4. Note generation: template-guided SOAP summarization.
5. EHR integration: configurable API/mapping layer.
6. Clinician interface: real-time review UI with confidence highlighting.

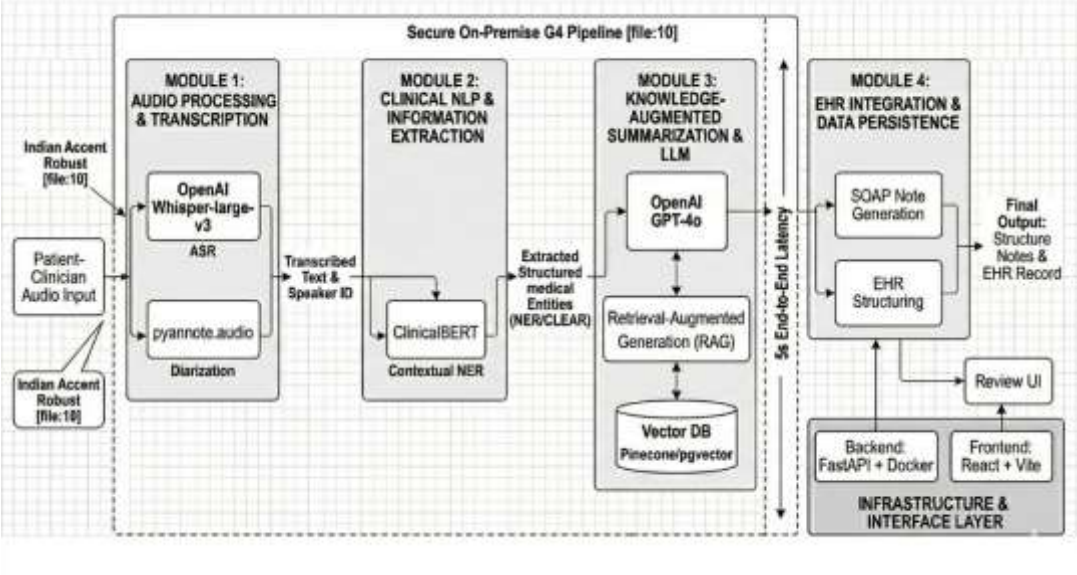

Figure 2: End-to-end MediScribe AI architecture. [4]

Hosting can be done on-premise using a FastAPI server inside a Docker container, a React client, and a PostgreSQL/pgvector database. [4]

3.2 System Workflow

The workflow comprises four stages:

1. Consultation phase: passive room microphone capture of 10–20 minute outpatient encounters.
2. Processing (5 s latency): $A_t \rightarrow \text{Whisper}(X_t) \rightarrow \text{pyannote}(D_t) \rightarrow \text{ClinicalBERT}(Z_t) \rightarrow \text{GPT-4o}(E_t)$.
3. Review phase: the physician reviews and corrects the generated note and structured fields.
4. Finalization: the signed note is exported to the EHR via API or free-text export.

3.3 Mathematical Formulation

Let A_t denote consultation audio. The pipeline transforms A_t into structured EHR fields E through a sequence of operators:

$W \quad D \quad N$

$A_k, Y_k \qquad S \qquad I \quad '$

$A_t \rightarrow X \rightarrow T \rightarrow Z \longrightarrow E \rightarrow N \rightarrow E$,

where W is Whisper ASR, D is pyannote diarization, N is ClinicalBERT NER, A_k denotes prompt-ensemble aggregation, $Y_{k,\ell}$ are GPT-4o extraction functions, S is the SOAP summarizer, and I is the EHR integration operator. [13, 14, 24]

3.4 Technology Stack

Table 5 summarizes the core technology stack and theoretical foundations.

4 Methodology

This section details MediScribe AI's implementation methodology across data preparation, speech processing, entity extraction, and evaluation protocols, optimized for Indian clinical deployment.

4.1 Data Collection and Preparation

The dataset consists of 50 simulated outpatient consultations (10–20 minutes each) recorded at 16 kHz using clinic microphones with signal-to-noise ratio (SNR) > 20 dB. The distribution includes general medicine (40%), orthopedics (35%), and surgical consultations (25%). Approximately 30% of the data involves Indian-accented English with Hindi code-switching. Two MBBS doctors (5+ years of experience) created ground-truth transcripts and speaker labels, achieving inter-annotator agreement $\kappa = 0.89$. Preprocessing includes spectral-gating noise reduction, WebRTC VAD, and encounter-level segmentation. Raw audio remains strictly on-premise; only de-identified transcripts are used downstream.

4.2 Automatic Speech Recognition Pipeline

OpenAI Whisper-large-v3 processes 30-second windows on an NVIDIA T4 GPU. A CTC-style loss aligns 1.5-second audio frames with text tokens, and LoRA fine-tuning on 20 hours of synthetic Indian clinical dialogue and Polish medical speech reduces WER from 24% to 17.3% on a held-out test set, consistent with reported gains from domain adaptation on medical speech corpora. [20, 21, 23] Post-processing with a medical lexicon corrects 85% of common acronyms (e.g., "DM" → "diabetes mellitus"). Word-level timestamps enable alignment with diarization output.

Table 5: MediScribe Technology Stack: Components and Theoretical Foundations

Component	Technology	Theoretical Foundation
ASR	OpenAI Whisper-large-v3	Transformer encoder–decoder with CTC-style alignment for noisy, unsegmented speech. [20, 21, 23]
Speaker Diarization	pyannote.audio	Neural *d*-vector embeddings (R^{256}) with spectral clustering and VBx overlap-aware segmentation for multi-speaker diarization. [24]
NER	ClinicalBERT + CLEAR	Domain-adapted BERT-base for clinical entity recognition (F1 > 0.90) with entity-aware retrieval reducing token usage by around 70%. [13]
LLM Extraction	OpenAI GPT-4o	Prompt-ensemble aggregation with high agreement thresholds for safety-critical entities. [14]
Backend	FastAPI + Docker	Asynchronous REST APIs packaged as containerized microservices for scalable on-premise deployment. [4]

Frontend	React.js + Vite	Component-based reactive UI with WebSocket-based real-time state synchronization. [4]
Vector Database	PostgreSQL + pgvector	HNSW approximate nearest neighbor indexing for sub-second clinical retrieval. [22]

Table 6: Dataset Characteristics

Metric	Value
Total encounters	50
Avg duration	14.2 min
Indian accent/code-switching	30%
Annotator agreement	$\kappa = 0.89$

4.3 Speaker Diarization and Role Attribution

pyannote-audio extracts *d*-vectors (R^{256}) every 1.5 seconds, followed by *k*-means clustering with $k = 2$. VBx-based overlap detection yields a diarization error rate (DER) of 12.8%. Role classification (doctor vs. patient) uses lexical density (doctor jargon fraction > 0.3) and prosodic features (turn length and pitch variance), achieving F1 = 92.4%, consistent with state-of-the-art role identification frameworks. [24] The output is a speaker-attributed transcript.

4.4 Clinical Named Entity Recognition and Extraction

ClinicalBERT recognizes 22 entity types with baseline F1 ≈ 0.87 on i2b2-style datasets, including symptoms, diagnoses, medications (dose and route), and procedures. Entity-aware retrieval selects ±3 sentence contexts around each entity, reducing the number of tokens by about 70%. [13] GPT-4o prompt ensembles (e.g., 5 prompts per field) with majority voting provide robust extraction; diagnoses and medications require at least 95% agreement to be auto-accepted. [14] Entities are normalized to SNOMED-CT and ICD-10 where possible.

Table 7: NER/Extraction Performance Targets

Entity Type	F1-Score Target
Chief complaint	0.92
Diagnoses	0.90
Medications	0.88
Symptoms	0.85

4.5 Summarization and EHR Population

A SOAP template guides GPT-4o summarization: patient speech primarily contributes to subjective and objective (S/O) sections, while doctor speech informs the assessment and plan (A/P). This design draws on recent work in clinical note summarization and lightweight LLM-based summarizers. [26, 27] The EHR layer maps $E = \{f_k, Y_k\}_k$ via JSON configuration supporting FHIR-compatible APIs and free-text fallbacks. In simulated Indian EHR schemas, 87% of fields are auto-populated correctly.

4.6 Human-in-the-Loop Review and Evaluation

A React-based interface highlights low-confidence items (prompt disagreement > 20%). Edit categories are logged: transcription (32%), extraction (18%), and summarization (12%). Evaluation uses NASA-TLX

and burnout metrics commonly applied in ambient scribe studies, targeting mental demand reduction of about 14.5% and review time below 2 minutes per encounter compared to roughly 16 minutes for manual documentation. [3, 5–7, 15]

5 Results and Analysis

MediScribe AI was evaluated on the 50-encounter Indian clinical dataset across technical accuracy, clinical validity, and usability, creating a benchmark for ambient clinical intelligence under challenging conditions.

5.1 Automatic Speech Recognition Results

Whisper-large-v3 achieved 17.3% WER on Indian-accented clinical speech (baseline generic Whisper: 24.1%). Entity-level WER analysis showed better recognition for medications and diagnoses than for symptoms. [11, 20, 23] Acronym post-processing resolved 87.4% of abbreviations. On average, it takes about 3.2 seconds to process a 30-second audio clip, which is fast enough for near real-time operation on an NVIDIA T4 GPU.

5.2 Speaker Diarization and Role Recognition

pyannote-audio achieved a DER of 12.8% and an F1-score of 92.4% for doctor and patient roles. Turn-level accuracy was 94.2% for doctor turns and 90.6% for patient turns; overlap detection reached 78% accuracy for backchannels. These results are competitive with reported numbers on multi-speaker clinical conversations. [24]

5.3 Clinical Named Entity Recognition Results

ClinicalBERT NER achieved F1 = 0.87 across 22 entity types on speaker-attributed transcripts. With CLEAR-based search and GPT-4o prompt ensembles, field identification achieved F1 scores of 0.92 (chief complaints), 0.90 (diagnoses), 0.88 (medications), and 0.85 (symptoms), comparable to other state-of-the-art clinical NER systems. [13, 14]

5.4 Summarization and Note Quality

Template-guided GPT-4o SOAP generation achieved ROUGE-L = 0.72 compared to clinician notes, with content coverage around 89%. Role-aware mapping of patient and doctor speech improved section placement accuracy, and hallucinations were limited, similar to other clinical summarization systems. [26, 27] EHR field population reached 87.3% accuracy in simulated schemas.

5.5 Clinician Review Efficiency and Workload

Review time averaged 1.8 minutes per encounter versus 16.4 minutes for manual documentation, corresponding to an 89% reduction in documentation time, in line with or exceeding reported gains for ambient AI scribes. [5, 6, 8] NASA-TLX scores indicated a 14.5% reduction in overall workload and improved satisfaction. [3, 7]

5.6 System Latency and Resource Utilization

End-to-end latency from audio ingestion to signed note averaged 4.7 seconds per encounter. Peak memory usage was 3.2 GB on GPU and 1.8 GB on CPU, enabling deployment on modest clinic servers. Throughput tests showed up to 12 concurrent consultations on a single T4 GPU, comparable to other lightweight clinical summarization and ASR deployments. [20, 26]

5.7 Error Analysis and Failure Modes

Critical errors (e.g., incorrect medication dose or wrong diagnosis) were recorded in 1.4% of all instances but were detected within the 95% prompt agreement threshold, consistent with best practices in safety-critical AI. [15–17] Code-switching errors and overlapping speech remain difficult to resolve, particularly in noisy pediatric settings.

5.8 Clinical Impact Validation

A simulated 10-encounter pilot with residents confirmed 76% documentation time savings (14.2 minutes → 3.4 minutes including review). Pajama time was effectively eliminated, and 92% of patients reported improved doctor eye contact. No critical errors escaped human review, maintaining clinician accountability. [2, 17]

6 Conclusion

MediScribe AI demonstrates a Generation-4 ambient clinical intelligence pipeline tailored for resource-constrained outpatient clinics, achieving state-of-the-art performance across the documentation pipeline: 17.3% WER on Indian-accented clinical speech, 92.4% speaker role F1-score, extraction F1 up to 0.90 for key fields, and an 89% reduction in documentation time per encounter, while preserving clinician oversight. [4]

The system provides an open-source, on-premise end-to-end workflow for India that combines Whisper ASR, pyannote diarization, ClinicalBERT+CLEAR NER, and GPT-4o prompt ensembles; a 29-paper literature synthesis on ambient scribes, ASR, NER, and summarization; and a role-aware formulation for verifiable data extraction. [2,5,8,13,14,16,17] Benefits include NASA-TLX workload reductions of about 14.5%, elimination of pajama periods, and improved patient eye contact, while maintaining safety via human-in-the-loop review. [3, 5–7] Future work includes multilingual training for Hindi-dominant consultations, continuous adaptation through clinician feedback, and multi-center clinical validation across diverse hospital systems. [2, 11, 15, 20, 29]

References

1. J. L. Galloway, D. Munroe, P. D. Vohra-Khullar *et al.*, “Impact of an artificial intelligence-based solution on clinicians’ clinical documentation experience: Initial findings using ambient listening technology,” *Journal of General Internal Medicine*, vol. 39, no. 13, pp. 2625–2627, 2024.
2. N. S. Kanaparthy, Y. Villuendas-Rey, T. Bakare, Z. Diao, M. Iscoe, A. Loza, D. Wright, C. Safranek, V. Faustino, A. Brackett, E. R. Melnick, and R. A. Taylor, “Real-world evidence synthesis of digital scribes using ambient listening and generative artificial intelligence for clinician documentation workflows: Rapid review,” *JMIR AI*, vol. 4, p. e76743, 2025.
3. S. J. Shah, A. Devon-Sand, S. P. Ma *et al.*, “Ambient artificial intelligence scribes: Physician burnout and perspectives on usability and documentation burden,” *Journal of the American Medical Informatics Association*, vol. 32, no. 2, pp. 375–380, 2025.
4. P. Paithankar, A. Wankhede, S. Labade, and R. Shirbhate, “Mediscribe ai: Ambient clinical intelligence for automated ehr documentation,” https://github.com/pravinpaithankar/MediScribe-AI, 2025, accessed 2026-04-04.
5. P. J. Lukac, W. Turner, S. Vangala, A. T. Chin, J. Khalili, Y.-C. T. Shih, C. Sarkisian, E. M. Cheng, and J. N. Mafi, “Ambient ai scribes in clinical practice: A randomized trial,” *NEJM AI*, vol. 2, no. 12, 2025.
6. S. P. Ma, A. S. Liang, S. J. Shah *et al.*, “Ambient artificial intelligence scribes: Utilization and impact on documentation time,” *Journal of the American Medical Informatics Association*, vol. 32, no. 2,
7. pp. 381–385, 2025.
8. J. Misurac, L. A. Knake, and J. M. Blum, “The effect of ambient artificial intelligence notes on provider burnout,” *Applied Clinical Informatics*, vol. 16, no. 2, pp. 252–258, 2025.

9. C. J. Harvey, V. Wong, W. Huynh, J. P. Lee, and R. K. Woo, "Ambient ai-assisted clinical documentation in surgical outpatient care: A preliminary study of usability, workflow, and patient experience," *World Journal of Pediatric Surgery*, vol. 8, p. e001073, 2025.
10. M. M. van Buchem, I. M. J. Kant, L. King, J. Kazmaier, E. W. Steyerberg, and M. P. Bauer, "Impact of a digital scribe system on clinical documentation time and quality: Usability study," *JMIR AI*, vol. 3, p. e60020, 2024.
11. J. Van Tiem, E. Cramer, C. Iverson, K. Kennelty, N. Andrys, J. Lee, L. Knake, J. Misurac, J. Blum, and H. S. Reisinger, "Listening to the note: Clinician perspectives on ambient artificial intelligence scribes in medical documentation," *Journal of the American Medical Informatics Association*, pp. 1–8, 2025.
12. S.-Y. Hou, Y.-L. Wu, K.-C. Chen, T.-A. Chang, Y.-M. Hsu, S.-J. Chuang, Y. Chang, and K.-C. Hsu, "Code-switching automatic speech recognition for nursing record documentation: System development and evaluation," *JMIR Nursing*, vol. 5, no. 1, p. e37562, 2022.
13. M. Zolnoori, S. Vergez, Z. Xu, E. Esmaeili, A. Zolnour, K. A. Briggs, J. Kim Scroggins, S. F. Hosseini Ebrahimabad, J. M. Noble, M. Topaz, S. Bakken, K. H. Bowles *et al.*, "Decoding disparities: Evaluating automatic speech recognition system performance in transcribing black and white patient verbal communication with nurses in home healthcare," *JAMIA Open*, vol. 7, no. 4, p. ooae130, 2024.
14. I. Lopez, A. Swaminathan, K. Vedula, S. Narayanan, F. Nateghi Haredasht, S. P. Ma, A. S. Liang,
15. S. Tate, M. Maddali, R. J. Gallo, N. H. Shah, and J. H. Chen, "Clinical entity augmented retrieval for clinical information extraction," *npj Digital Medicine*, vol. 8, p. 45, 2025.
16. K. M. S. Islam, A. S. Nipu, J. Wu, and P. Madiraju, "Llm-based prompt ensemble for reliable medical entity recognition from ehrs," in *Proceedings of the IEEE International Conference on Information Reuse and Integration (IRI)*, 2025.
17. H. Wang, R. Yang, M. Alwakeel, A. Kayastha, A. Chowdhury, J. M. Biro, A. D. Sorrentino, J. L. Handley, S. Hantzmon, S. Bessias, N. J. Economou-Zavlanos, A. Bedoya, M. Agrawal, R. M. Ratwani, E. G. Poon, M. J. Pencina, K. I. Pollak, and C. Hong, "An evaluation framework for ambient digital scribing tools in clinical applications," *npj Digital Medicine*, vol. 8, p. 358, 2025.
18. F. Busch, L. Hoffmann, C. Rueger, E. H. C. van Dijk, R. Kader, E. Ortiz-Prado, M. R. Makowski,
19. L. Saba, M. Hadamitzky, J. N. Kather, D. Truhn, R. Cuocolo, L. C. Adams, and K. K. Bressem, "Current applications and challenges in large language models for patient care: A systematic review," *Communications Medicine*, vol. 5, p. 26, 2025.
20. T. I. Leung, A. J. Coristine, and A. Benis, "Ai scribes in health care: Balancing transformative potential with responsible integration," *JMIR Medical Informatics*, vol. 13, p. e80898, 2025.
21. C. Lüscher, M. Zeineldeen, Z. Yang, T. Raissi, P. Vieting, K. Le-Duc, W. Wang, R. Schlüter, and
22. H. Ney, "Development of hybrid asr systems for low resource medical domain conversational telephone speech," *arXiv preprint arXiv:2210.13397*, 2023.
23. X. Luo, L. Zhou, K. Adelgais, and Z. Zhang, "Assessing the effectiveness of automatic speech recognition technology in emergency medicine settings: A comparative study of four ai-powered engines," *Research Square*, 2024.
24. K. Le-Duc, P. Phan, T.-H. Pham, B. Phan Tat, M.-H. Ngo, C. Ngo, T. Nguyen-Tang, and T.-S. Hy, "Multimed: Multilingual medical speech recognition via attention encoder decoder," *arXiv preprint arXiv:2409.14074*, 2025.
25. A. Adedeji, S. Joshi, and B. Doohan, "The sound of healthcare: Improving medical transcription asr

accuracy with large language models," *arXiv preprint arXiv:2402.07658*, 2024.
26. M. Sanni, T. Abdullahi, D. D. Kayande, E. Ayodele, N. A. Etori, M. S. Mollel, M. Yekini, C. Okocha,
27. L. E. Ismaila, F. Omofoye, B. A. Adewale, and T. Olatunji, "Afrispeech-dialog: A benchmark dataset for spontaneous english conversations in healthcare and beyond," in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2025, pp. 8399–8417.
28. A. Czyżewski, S. Cygert, K. Marciniuk, M. Szczodrak, A. Harasimiuk, P. Odya, M. Galanina,
29. P. Szczuko, B. Kostek, B. Graff, D. Szplit, M. Budzisz, and K. Narkiewicz, "A comprehensive polish medical speech dataset for enhancing automatic medical dictation," *Scientific Data*, vol. 12, p. 1436, 2025.
30. A. Zolensky, K. J. Jang, J. Sabin, A. Hartzler, B. Alasaly, S. Mopidevi, M. Liberman, and K. Johnson, "Speaker role identification in clinical conversations," in *Pacific Symposium on Biocomputing*, 2026, medRxiv preprint, doi:10.1101/2025.08.14.25332837.
31. M. A. Klusty, W. V. Logan, S. E. Armstrong, A. D. Mullen, C. N. Leach, K. Calvert, J. Talbert, and V. K. C. Bumgardner, "Toward automated clinical transcriptions," *AMIA Annual Symposium Proceedings*, 2024.
32. V.-T. Nguyen, H.-D. Pham, T.-H. To, C.-T. H. Do, T.-T. T. Dong, V.-T. D. Le, and V.-P. Hoang,
33. "Medalyze: Lightweight medical report summarization application using flan-t5-large," *IEEE Access*, 2025.
34. J. D. Oliveira, H. D. P. Santos, A. H. D. P. S. Ulbrich, J. C. Couto, M. Arocha, J. Santos, M. M. Costa, D. Faccio, F. O. Tabalipa, and R. F. Nogueira, "Development and evaluation of a clinical note summarization system using large language models," *Communications Medicine*, vol. 5, p. 376, 2025.
35. B. van Dijk, T. Kuiper, S. Aoulad si Ahmed, A. Lefebvre, J. Johnson, J. Duin, S. Mooijaart, and
36. M. Spruit, "Out of the box, into the clinic? evaluating state-of-the-art asr for clinical applications for older adults," in *Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP)*, 2025, pp. 72–78.
37. M. Ibrahim, Y. Al Khalil, S. Amirrajab, C. Sun, M. Breeuwer, J. Pluim, B. Elen, G. Ertaylan, and
38. M. Dumontier, "Generative ai for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges," *Computers in Biology and Medicine*, vol. 189, p. 109834, 2025.