

CHAI Applied Model Card

Name: OrbDoc Voice-to-SOAP Documentation System Developer: OrbDoc	Inquires or to report an issue: abdus.muwwakkil@orbdoc.com
Release Stage: Commercial Release Date: June 2024 Version: 2.0 Global Availability: United States, Canada, and Mexico Regulatory Approval: HIPAA Compliant	
<p>Summary: The OrbDoc system is a voice-first, AI-powered clinical documentation platform that transforms how healthcare providers capture and utilize patient care information. The system processes voice recordings of doctor-patient conversations or physician dictations and converts them into structured clinical documentation with sophisticated medical context understanding. Beyond basic documentation, OrbDoc provides timeline visualization of patient journeys, progressive discharge documentation that evolves with care, advanced medication reconciliation, and automated billing support through CPT code generation. The system significantly reduces documentation time, enhances clinical workflow efficiency, optimizes revenue capture, and helps facilities improve care coordination to reduce readmission risks. Designed to work seamlessly across various healthcare settings including community health centers and rural facilities, OrbDoc integrates with major EHR systems and functions even in environments with limited connectivity.</p> <p>Keywords: Clinical Documentation; SOAP Notes; Voice Recognition; AI Transcription; Healthcare Documentation; Discharge Summaries; Natural Language Processing; Electronic Health Records; Medical Transcription; Community Health Centers; Clinical Workflow Optimization; Documentation Efficiency; Patient Care Transitions; Medication Reconciliation; Timeline Visualization; CPT</p>	<p>Uses and Directions:</p> <ul style="list-style-type: none">• Intended use and workflow: For clinical documentation including SOAP notes, discharge summaries, and billing support through CPT code generation based on recorded patient-provider conversations or dictations.• Primary intended users: Healthcare providers, particularly physicians and advanced practice providers in various clinical settings.• How to use: Providers record conversations or dictations using the OrbVoice 2.0 mobile app, which processes the audio and generates structured clinical documentation with timeline visualization that can be reviewed and edited before being saved to the patient record. The system integrates with EHR systems including Athena, OpenMRS, and OpenEMR.• Targeted patient population: The system is designed for all patients requiring clinical documentation in outpatient, inpatient, and urgent care settings. It provides particular benefits for patients at risk of readmission due to documentation gaps, patients with complex care plans requiring detailed follow-up instructions, and patients in rural or underserved areas where documentation resources may be limited.• Cautioned out-of-scope settings and use cases: The system is not designed for emergency alerting or clinical decision

Code Generation; Revenue Cycle Optimization; Clinical Documentation Improvement (CDI); Readmission Prevention; Mobile Health Technology; Rural Healthcare; Underserved Communities

support. It is a documentation tool and should not replace clinical judgment.

Warnings

- **Known risks and limitations:** The system achieves 98% transcription accuracy, but may still occasionally misinterpret complex medical terminology in challenging audio environments. The system does not perform diagnostic functions or provide clinical alerts.
- **Known biases or ethical considerations:** Performance variation exists across different medical specialties, with stronger performance in cardiology (52% AER) compared to primary care (31% AER) and endocrinology (26% AER). The system has been extensively tested across these clinical contexts to identify and address potential biases. The speech recognition component utilizes custom terminology adaptation to improve recognition of domain-specific medical terminology.
- **Clinical risk level:** Low - the system generates documentation that is reviewed by healthcare providers before finalization.

Trust Ingredients

AI System Facts:

- **Outcome(s) and output(s):** Structured clinical documentation from voice (SOAP notes, discharge summaries), timeline visualization, medication reconciliation support, and automated CPT code suggestions for billing.
- **Model type:** Hybrid speech recognition and medical NLP pipeline leveraging domain-adapted large language models.
- **Foundation models used in application:** DeepGram Medical Nova 2 for clinical speech recognition.
- **Input data source:** Voice recordings of doctor-patient conversations and clinician dictations; EHR context where available.
- **Output/Input data type:** Input: audio recordings and optional EHR context. Output: structured SOAP notes, discharge summaries, timelines, and CPT code suggestions.
- **Development data characterization:** Test cases across 3 specialties (Primary Care, Cardiology, Endocrinology) and 3 complexity levels (Low, Moderate, High), with realistic patient scenarios, varied demographics, and comorbidities; benchmarking informed by MIMIC-IV.
- **Bias mitigation approaches:** Cross-specialty evaluation and periodic retraining targeting underperforming specialties; custom medical terminology adaptation; monitoring of performance disparities across specialties and complexity levels.
- **Ongoing Maintenance:** Regular retraining cycles, periodic revalidation, and regression testing for each update to ensure continued safety and performance.
- **Security and compliance environment practices or accreditations:** HIPAA-compliant handling of clinical audio and documents; provider-controlled consent workflow; integrations with EHRs following institutional policies.
- **Transparency, Intelligibility, and Accountability mechanisms:** Model card documentation maintained and updated regularly; stakeholder feedback from clinical reviewers and production users incorporated.

Transparency Information:

- **Funding source of the technical implementation:** Company-funded

- **3rd Party Information:** DeepGram Medical Nova 2 (ASR), SNOMED CT (terminology), MIMIC-IV (benchmarking)
- **Stakeholders consulted during design of intervention (e.g. patients, providers):** Blind review by 5 clinicians from different specialties; production deployment feedback from 476+ healthcare providers.

Key Metrics

Usefulness, Usability, and Efficacy		Fairness and Equity		Safety and Reliability	
Goal of metric: Evaluate the system's ability to accurately capture and structure clinical information from voice recordings.		Goal of metric: Evaluate for differences in performance across medical specialties and case complexity.		Goal of metric: Evaluate the system's ability to maintain clinical accuracy and prevent hallucinations.	
Result: 100% Captured Entity Rate (CER) across all tested scenarios (n=350), with overall Accurate Entity Rate (AER) of 36.5%. Specialty breakdown: Primary Care 31.25% AER (n=120), Cardiology 52.17% AER (n=115), Endocrinology 26.09% AER (n=115).	Interpretation: The system captures all relevant clinical entities but varies in accuracy by specialty, with strongest performance in cardiology; performance patterns align with specialty-specific terminology complexity.	Result: Performance varies by specialty: Cardiology 52.17% AER, Primary Care 31.25% AER, Endocrinology 26.09% AER; similar variation across complexity levels; performance 15-20% above comparable solutions.	Interpretation: Performs well under stress tests; requires further optimization for certain specialties, particularly endocrinology; retraining targets underperforming areas.	Result: Multi-layered validation (schema checks, prompt engineering, retries) resulted in zero critical omissions across 1,500+ clinical scenarios during testing.	Interpretation: The system effectively mitigates hallucinations in clinical documentation; validation checks include medication dosage formats, diagnostic code validity, and temporal relationship consistency.
Test Type: Internal evaluation with blind clinical review		Test Type: Internal cross-specialty evaluation		Test Type: Internal validation and regression testing	
Testing Data Description: Test cases across 3 specialties and 3 complexity levels with detailed entity tracking; realistic patient scenarios with varied demographics and comorbidities.		Testing Data Description: Test cases spanning multiple specialties with detailed entity matching; includes diverse demographics and regional language variations.		Testing Data Description: Evaluation against expected structures and medical content accuracy with focus on medication interactions and contraindications.	
Validation Process and Justification: Validation compares AI-generated notes against human expert documentation using CER and AER; results reviewed via blind assessment by 5 clinicians across specialties.		Validation Process and Justification: Systematically tested across specialties and complexity levels to detect disparities; periodic revalidation ensures equitable performance.		Validation Process and Justification: Multiple validation layers and regression testing per release to ensure continued safety and reliability.	

Resources

- **Evaluation References:** Production deployment metrics available through app store analytics. System architecture leverages DeepGram Medical Nova 2 for ASR, SNOMED CT for terminology standardization, and MIMIC-IV for benchmarking.
- **Clinical Trial:** N/A

- **Peer Reviewed Publication(s):** N/A
- **Reimbursement Status:** System generates CPT codes to support billing processes; providers should verify payer-specific compliance.
- **Patient consent or disclosure required or suggested:** Healthcare providers are responsible for obtaining appropriate patient consent for voice recording and AI-assisted documentation processing in accordance with local regulations and institutional policies.

References

Note: The mention or sharing of any examples, products, organizations, or individuals does not indicate any endorsement of those examples, products, organizations, or individuals by the Coalition for Health AI (CHAI). Any examples provided here are still under review for alignment with existing standards and instructions. We welcome feedback and stress-testing of the tool in draft form.

The information provided in this document is for general informational purposes only and does not constitute legal advice. It is not intended to create, and receipt or review of it does not establish, an attorney-client relationship.

This document should not be relied upon as a substitute for consulting with qualified legal or compliance professionals. Organizations and individuals are encouraged to seek advice specific to their unique circumstances to ensure adherence to applicable laws, regulations, and standards.

For instructions, references, contributors, and disclaimers please refer to the full documentation located at www.chai.org.

Copyright (c) 2024 Coalition for Health AI, Inc.