



WHITE PAPER

ClinixSummary Architecture: A Technical Overview

System architecture, model pipeline, audio processing, NLP layers, and output generation. A comprehensive look at how ClinixSummary transforms clinical audio into structured documentation.

Published:	March 2025
Document ID:	WP-2025-001
Classification:	Public
Author:	ClinixSummary Research Team

This document is provided for informational purposes only. ClinixSummary is a product of GATMEDI. The information contained herein is proprietary and confidential. Reproduction or distribution without prior written consent is prohibited. For the latest version, visit clinixsummary.ai.



1. Executive Summary

ClinixSummary is a vertically integrated AI platform purpose-built for clinical documentation. Unlike generic transcription services that bolt medical vocabulary onto consumer speech-to-text engines, ClinixSummary employs a proprietary multi-stage pipeline that transforms raw clinical audio into structured, specialty-aware clinical notes in real time.

This white paper provides a technical overview of the ClinixSummary architecture for CTOs, IT leaders, and technical evaluators. It covers the end-to-end data flow from audio capture through to final document generation, including the intermediate NLP stages that distinguish ClinixSummary from commodity transcription services.

2. System Architecture Overview

The ClinixSummary platform comprises five primary subsystems, each designed as an independently scalable microservice communicating over encrypted internal channels:

- **Audio Ingestion Layer** — Handles real-time audio capture, buffering, and preprocessing. Supports both ambient (multi-speaker) and single-speaker dictation modes with automatic detection.
- **Speech Recognition Engine** — Proprietary ASR models fine-tuned on medical speech corpora spanning 40+ specialties. Achieves word error rates (WER) significantly below commercial general-purpose engines on clinical vocabulary.
- **Clinical NLP Pipeline** — Multi-stage processing including speaker diarisation, medical named entity recognition (NER), section classification, and clinical context inference.
- **Document Assembly Engine** — Transforms structured NLP outputs into specialty-appropriate clinical note formats (SOAP, BIRADS, operative reports, therapy notes, etc.).
- **Integration & Delivery Layer** — FHIR-compatible APIs, direct EHR/EMR export, and the ClinixSummary Console for review and editing.

3. Audio Ingestion & Preprocessing

3.1 Capture Modes

ClinixSummary supports three distinct capture modes, automatically detected by the Auto Ambient Dictate system:

- **Ambient Mode** — Multi-speaker clinical consultation captured via device microphone. Speaker diarisation separates clinician from patient dialogue.
- **Dictation Mode** — Single-speaker post-visit narration. Optimised for higher speech rates and medical shorthand.
- **Operative Mode** — Real-time narration during surgical procedures. Handles background noise, team communication, and procedural terminology.

3.2 Signal Processing

Raw audio undergoes adaptive noise reduction, automatic gain control, and voice activity detection (VAD) before reaching the ASR engine. The preprocessing pipeline runs on-device where possible to minimise data transmission and



latency, with server-side processing as fallback.

4. Speech Recognition Engine

ClinixSummary's ASR is not a fine-tuned wrapper around a third-party engine. It is a proprietary encoder-decoder architecture trained from the ground up on de-identified clinical speech data. Key characteristics include:

- Vocabulary coverage of 250,000+ medical terms across ICD-10, CPT, SNOMED-CT, and specialty-specific lexicons.
- Real-time streaming inference with <500ms latency from speech to preliminary transcript.
- Multi-language support (English, French, Spanish, Portuguese, Italian, Arabic) with in-consultation code-switching capability.
- Accent and dialect robustness trained on geographically diverse clinical speech samples.

5. Clinical NLP Pipeline

5.1 Speaker Diarisation

In ambient mode, speaker diarisation identifies and separates clinician speech from patient speech, supporting up to 4 concurrent speakers. This is critical for accurate attribution of reported symptoms (patient) versus clinical assessments (clinician).

5.2 Medical Named Entity Recognition

A domain-specific NER model extracts clinical entities including diagnoses, medications, dosages, procedures, anatomical references, lab values, and temporal markers. Entity extraction feeds downstream section classification and coding suggestion modules.

5.3 Section Classification

A classification model assigns transcript segments to appropriate clinical note sections (Chief Complaint, History of Present Illness, Review of Systems, Physical Examination, Assessment, Plan, etc.). Section schemas are specialty-configurable.

5.4 Clinical Context Inference

Beyond entity extraction, ClinixSummary's contextual reasoning layer infers unstated clinical logic — for example, understanding that a discussion of 'the usual dose' refers to a previously established medication regimen, or that 'better than last time' implies comparison to a prior visit.

6. Document Assembly

The Document Assembly Engine takes structured NLP outputs and generates a complete clinical note in the appropriate format for the specialty and context. Templates are maintained per specialty and are continuously refined through the Kai-zen feedback loop.

Output formats include SOAP notes, procedure notes, therapy session summaries, dental charting narratives,



psychiatric evaluations, veterinary SOAP, and free-form dictation transcripts. Each format enforces completeness checks — flagging potential omissions before the clinician finalises the note.

7. Infrastructure & Scalability

ClinixSummary operates on a cloud-native infrastructure with the following characteristics:

- Auto-scaling compute clusters for ASR and NLP inference, handling variable load without performance degradation.
- Geographic data residency options (US, EU, UK, AU) to meet jurisdictional compliance requirements.
- 99.9% uptime SLA with multi-region failover.
- All inter-service communication encrypted with TLS 1.3; data at rest encrypted with AES-256.

8. Conclusion

ClinixSummary's architecture reflects a fundamental design principle: clinical documentation is a clinical problem, not a transcription problem. Every layer of the stack — from audio capture to document assembly — is purpose-built for the medical domain. This vertical integration enables accuracy, speed, and specialty awareness that cannot be achieved by layering medical prompts onto generic AI services.