CLAI Service Definition Document

About CLAI

CLAI is a comprehensive ambient scribe platform built for NHS workflows. Our rapidly growing clinical user base can generate letters, summaries and clinical notes directly from audio recordings, automatically extract a range of clinical codes and create follow up orders – all directly integrated into EPR via a truly frictionless workflow automation.

CLAI are fully compliant with all required NHS certifications, including an MHRA class 1 registration.

Our end-to-end workflow integration is saving up to 30% of clinician time in clinic, this equates to as many as 3 extra patient encounters within a standard clinic template.

CLAI Functionality Overview

- **EHR integration:** Launch with full patient context directly from your EPR. Full write back and downstream workflow integration for documents, clinical codes and follow up orders. Push patient letters directly to the NHS App, in near real time.
- **Oracle Cerner MPage support:** Publish your own dedicated m-page for Ambient Scribe or use our flexible MPage component to embed CLAI directly within your existing workflow.
- **Full RBAC support within your EPR:** Directly manage user access directly from Oracle Cerner, with full control over user groups, roles and session tracking.
- **Letter templates:** Full control and endlessly customisable options for letter templates. Choose from a bank of pre-defined options or simply create your own.
- **Clinical coding:** Automatic generation and EPR write back of ICD-10, SNOMED-CT, OPCS and RTT codes.
- **Follow up orders:** Follow up orders are directly generated from your consultation transcript and written to EPR, saving minutes of screen time per consultation.

- **Session continuity across devices:** Begin recording on your smartphone and seamlessly transfer your session to a desktop, or vice versa. Pause and resume consultations. Progress is saved securely at every step.
- **Comprehensive analytics:** A full suite of performance and usage metrics are available in real time dashboards. Track letter generation, scribe accuracy, edit rates and completion times.
- **Safe & secure by design:** Full clinical oversight and control, before submission to EHR. Completed sessions are deleted and data is fully encrypted in transit and at rest. Fully compliant with all necessary NHS security certifications, including a Class I MHRA registration.

Clinical System Integration

CLAI provides an expert and fully managed systems integration service, with a core specialty in the Oracle Cerner EPR platform. Using either HL7 V2.x or FHIR R4 (where supported), our integration is enabled by a dedicated integration engine.

Additional systems are also directly supported. Please contact us to learn more.

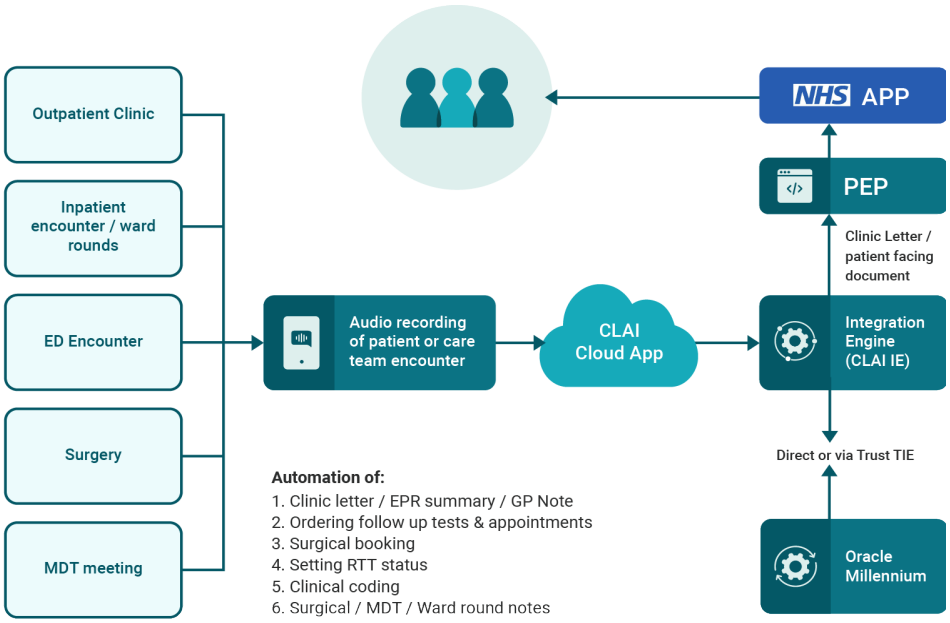

Fig. 1. CLAI top level system architecture

System Accuracy

The CLAI platform offers industry leading accuracy levels. Our platform output is continuously monitored across a wide range of accuracy metrics, including a regular and rigorous third-party clinical review cycle.

Key metric results include:

Testing Metrics	
Precision	79%
Recall	99%
WER	8%
User Reported Feedback	
User Reported Feedback on Hallucination in the Letter/Note	1%
User Reported Acceptance Rates	
No edits required	74%
Minor edits required	20%
Major edits required	6%

Fig. 2. System accuracy metrics

Information Governance and Data Protection

Model Governance

CLAI utilises commercially available LLM platforms and does not train or fine-tune foundational models on patient data. However, to ensure safety and accuracy, we maintain rigorous governance over our Evaluation and Testing Datasets (used for prompt engineering and accuracy benchmarking). Our procedures align with the NHS Good Practice Guidelines for AI and the EU AI Act..

Data Capture & Minimisation

We strictly adhere to data minimisation by using only specific, relevant datasets required for validation:

- **Sample/Actor Audio:** Produced by UK NHS clinicians role-playing scenarios containing no real patient data.
- **Anonymised Transcripts:** Text-based transcripts where all PII is removed at the source. We do not store or process the corresponding audio or clinical artefacts (e.g., scan images), ensuring we hold only the minimum text required to validate clinical accuracy.

Data Lineage & Provenance

We maintain a documented Data Lineage Log for all evaluation data.

- **Source:** Any real-world data originates solely from:
    - Clinician-patient discussions where the patient has explicitly consented to recording and where no reference to their name occurs in the audio file; OR
    - CQC-registered NHS providers under an active Data Processing Impact Assessment (DPIA) and Data Processing Agreement (DPA).
- **Traceability:** Every data point is tagged with its origin, version, and the specific suppression logic applied. This satisfies EU AI Act Article 10(3) regarding appropriate data collection and provenance documentation.

Anonymisation Approach

- **Process:** Consultation transcripts are processed via a UK/EEA-hosted, ISO 27001 certified data processor (Microsoft Azure and/or Google Gemini).
- **Technique:** PII is systemically redacted at the point of origin using Named Entity Recognition (NER) before ingestion into our test environment. This is irreversible anonymisation, not pseudonymisation; the resulting data cannot be programmatically re-identified.
- **Re-identification Mitigation:** To mitigate deductive disclosure risks in small samples:
    - We apply "Outlier Suppression": Transcripts containing ultra-rare clinical conditions that could theoretically identify a patient are excluded from the test set, via human review.
    - We adhere to ICO Anonymisation Code of Practice guidelines on "motivated intruder" tests.

Bias Assessment

We actively engineer our test datasets to detect and mitigate performance disparities (fairness testing), aligning with EU AI Act Article 10(3-4):

- **Demographic Coverage:** We deliberately sample audio/transcripts across diverse demographics (gender, regional accents, non-native English speakers) and clinical specialties.
- **Benchmarking:** We measure Word Error Rate (WER) and Hallucination Rate disaggregated by these groups. If a disparity is found, we update our system prompts and re-test before release.
- **Sampling:** For anonymised consultations, we use Stratified Random Sampling to ensure representation across different medical specialties (e.g., Cardiology vs. Paediatrics).

Alignment to ICO Guidance

- **Fairness:** We test for statistical accuracy parity across groups to prevent discriminatory outcomes.
- **Transparency:** Our privacy notices explicitly detail how evaluation data is used.
- **Accountability:** We maintain an auditable record of all testing outcomes and data protection checks.

EU AI Act – High-Risk Alignment

Although currently acting as a tool under clinician supervision, we align with High-Risk AI System requirements:

- **Data Governance (Art 10):** Our datasets are relevant, representative, and free of errors.
- **Technical Documentation (Art 11/12):** We maintain detailed records of data provenance and validation metrics.
- **Transparency (Art 13):** We provide clear instructions to users regarding the system's capabilities and the nature of the data it was evaluated on.

Accreditations and Standards Compliance

DTAC	Current and approved
DSPT	Current, standards met
Cyber Essentials Plus	Current, passed November 2025
ISO 27001	Current, Mayden House Ltd
Crest approved Pen Test	Performed at least annually and after any major infrastructure updates – last test 11/2025. No critical or high issues at last test and 2 minor issues fully remediated.
MHRA Class 1	#34230 – registered 24/02/2025
GDPR	Fully compliant

Fig. 3. CLAI accreditations and standards

Service Performance and Availability

CLAI delivers validated, real-time performance tailored for acute clinical environments, ensuring Zero-Lag interaction. CLAI provides its service at scale with high responsiveness.

- Architecture, Initiation and Latency: We utilise a decoupled, cloud-native architecture. The frontend is delivered via Azure Static Web Apps behind Azure Front Door, ensuring the UI remains responsive and stateful (latency <50ms) independently of backend processing load. Transcription initiation occurs in <500ms. Crucially, we do not use fragile WebSockets; audio is buffered locally and submitted via resilient API calls, ensuring stability even on potentially intermittent hospital Wi-Fi.
- Load Testing: recent stress testing simulated 180 concurrent active audio streams in one minute, equating to 10,800 transcriptions per hour and therefore able to support well in excess of ~5,000 active users and can easily be scaled further. The system maintained 100% persistence and zero dropped connections, with CPU usage managed via Azure App Service Auto scale.

We provide an "Always-On" service architecture designed for mission-critical healthcare use.

- **Availability:** We have a proven ≥99.7% uptime (SLA) over the last 12 months. Critical services (Database, API) utilise Zone-Redundant Storage (ZRS), synchronously replicating data across physically separated availability zones to ensure zero data loss (RPO = 0).
- **Disaster Recovery (DR):** In a catastrophic regional outage, automated health probes trigger traffic re-routing to our secondary region. Validated failover tests confirm a Recovery Time Objective (RTO) of just 43 seconds (App Service warm-up time), significantly outperforming the ≤5-minute requirement.
- **Non-Disruptive Updates:** We utilise Blue/Green deployment slots. New versions are deployed and health-checked in parallel before traffic is instantly swapped, resulting in zero downtime for users during updates.

CLAIs architecture is designed to deliver a continuous uninterrupted service to users.

Support and Service Levels

We provide a comprehensive support wrap integrated with the Trust's operational flow.

- **Coverage:** 24/7/365 coverage for Priority 1 (Critical) incidents.
- **Response**: Guaranteed triage within 15 minutes for P1 issues, with a target resolution time of <4 hours.
- **Documentation:** Monthly Service Level Reports are provided to the Trust, detailing uptime, incident volume, and resolution performance. SLA performance dashboards are available on demand.
- **Escalation:** Automated escalation paths ensure direct access to the On-Call Engineering Lead if SLAs are at risk.

Case Study (Critical Incident Drill): To validate our resolution capability, we conducted a "Total Region Failure" simulation. Upon simulated loss of the UK South region, our automated failover logic restored full-service availability in just 43 seconds. This exercise supports our ability to resolve catastrophic P1 incidents significantly faster than the 4-hour target.

Category	Response Time	Resolution Time	Response window	Description
P1	1 hour	4 hours	24/7/365	**Business critical failure.** A critical incident that results in major or total loss of platform functionality, preventing the use of core features, affecting all or most users.
P2	4 hour	8 hours	0830-1700, Mon - Friday	**System defect with no immediate workaround.** An error in the platform, which may hinder useful operation but does not prevent the use of core features. No immediate workaround is available.
P3	4 hour	3 business days	0830-1700, Mon - Friday	**System defect with an immediate workaround.** An error in the platform, which may hinder useful operation but does not prevent the use of core features. A work around exists and is deployed within 4 hours.
P4	1 business day	5 business days	0830-1700, Mon - Friday	**Minor error.** An isolated or minor error that does not significantly affect functionality but may reduce overall intended system efficiency.

Fig. 4. CLAI service levels and support incident classifications

Customer Onboarding and Offboarding

Service Design and Workflow Configuration

During the project mobilisation phase CLAI will work directly with you to understand your specific project goals and requirements. Typically this phase is concluded within 4 weeks of project commencement.

The CLAI platform supports a high degree of customisation – this is especially relevant for EPR driven workflows where a high degree of variation exists, even across providers running the same EPR. Significant benefit can be realised from updating or outright replacing existing legacy workflows (e.g. patient checkout).

Deployment and Mobilisation

Building on the outputs of the service design phase, CLAI will work with you to develop and publish a comprehensive project plan, governance model and delivery roadmap. CLAI will assign dedicated project management and technical delivery resource to ensure a successful project outcome.

You will have direct access to CLAI subject matter experts ('SMEs') and we will work together to identify CLAI champions to enable on the ground adoption support.

Training

CLAI operates a train the trainer model. We provide both in person and remote training, supporting key 'super users' and trainers with access to training materials and live training sessions.

CLAI conducts regular training reviews and provides a ticket based support system to capture issues or product enhancement suggestions.

Offboarding

The CLAI platform does not persist patient data. PII is held securely only until an encounter is concluded and submitted to EPR.

The majority of user access and audit log data is held within EPR.

Anonymised platform use data and reports will be transferred to the customer during the offboarding process. CLAI will securely store this data for up to one year post contract termination, to allow ample time for data transfer to be actioned.

Pricing, Ordering and Invoicing

Options for procuring and configuring CLAI are set out in Mayden's pricing document which is provided on the Digital Marketplace.

CLAI is provided as a SaaS platform, with costs directly determined by the scale of a proposed deployment. Price breaks offer greater value as service volumes increase.

For whole provider deployments CLAI offers enterprise pricing with no usage caps or additional volume based charges.

Project costs are split between one off implementation charges and ongoing recurring license fees.

Ongoing license fees are billed quarterly in advance.