



// METHODOLOGY DOCUMENT

# AI Model Evaluation Methodology

VERSION 1.0 – JUNE 2026

This document describes the evaluation methodology used by ClearanceAI to assess AI models deployed in regulated industries. It defines the three evaluation layers, the scoring framework, regulatory framework mapping, data handling standards, and the basis for the ClearanceAI Certified designation. This methodology is publicly available for inspection by clients, regulators, and independent reviewers.

● PUBLICLY INSPECTABLE

● INDEPENDENTLY REVIEWABLE

● NAMED CLAUSE MAPPING

● SHA-256 EVIDENCE ANCHORED

● ZERO CLIENT DATA RETENTION

● NO MODEL TRAINING USE

// CONTENTS

# Table of Contents

## SECTION 1

Introduction and Purpose PAGE 3

---

Why Independent Evaluation Matters PAGE 3

---

Scope of This Methodology PAGE 3

---

## SECTION 2

Regulatory Frameworks Evaluated Against PAGE 4

---

Industry Coverage PAGE 4

---

## SECTION 3

The Three Evaluation Layers PAGE 5

---

Layer 01 — Automated Evaluation Engine PAGE 5

---

Layer 02 — Credentialed Expert Review PAGE 5

---

Layer 03 — Compliance Assessment Report PAGE 5

---

## SECTION 4

Scoring Methodology PAGE 6

---

Verdict States PAGE 6

---

ClearanceAI Certified Designation PAGE 6

---

## SECTION 5

Data Handling and IP Protection PAGE 7

---

Zero-Retention Policy PAGE 7

---

No Model Training Use PAGE 7

---

Limitations and Disclaimer

PAGE 8

What a ClearanceAI Report Does Not Certify

PAGE 8

Document Control

PAGE 8

---

// SECTION 01

## Introduction and Purpose

ClearanceAI is an independent AI model evaluation platform serving regulated industries — healthcare, financial services, defense, and energy. This document describes the evaluation methodology ClearanceAI applies when assessing AI models against applicable regulatory frameworks and delivering a formal Compliance Assessment Report.

This methodology is published publicly so that clients, regulators, enterprise procurement teams, and independent reviewers can inspect the basis on which ClearanceAI produces its evaluation conclusions. Transparency in methodology is a prerequisite for credibility in independent assessment.

---

### Why Independent Evaluation Matters

AI models deployed in regulated industries are subject to increasingly specific requirements — FDA 2025 AI Guidance, NIST AI RMF 1.0, ISO 42001, CMS non-discrimination requirements, and DoD AI Ethics Principles, among others. These frameworks require documented evidence of compliance, not merely internal assertions.

Internal governance tools — however well designed — are self-reported by definition. They document what an organization believes about its own AI model. When a hospital's legal team, a bank's risk committee, or a government procurement officer asks for independent validation, a self-assessment does not satisfy the question. The same principle has governed financial reporting for over a century: internal controls are necessary, but external audit is required for third-party reliance.

**ClearanceAI's role** is to provide the independent assessment layer that internal tools cannot — a formal third-party evaluation, produced by a credentialed external party, that regulators, buyers, investors, and boards can rely on independent of the AI vendor's own representations.

---

### Scope of This Methodology

This methodology applies to all ClearanceAI evaluations conducted under a signed Master Services Agreement and Statement of Work. It covers:

- AI models and AI-enabled software tools deployed or intended for deployment in regulated industries
- Software as a Medical Device (SaMD) under FDA jurisdiction

- AI tools subject to CMS non-discrimination and prior authorization requirements
- Financial AI models subject to SR 11-7 model risk guidance
- Defense AI systems subject to DoD AI Ethics Principles and CDAO framework
- Any AI system where regulatory compliance documentation is required for market access or enterprise procurement

This methodology does not cover post-deployment operational monitoring. Ongoing monitoring is addressed through ClearanceAI's monthly Smoke Test service, described separately.

---

// SECTION 02

## Regulatory Frameworks Evaluated Against

ClearanceAI evaluates AI models against the specific frameworks applicable to the client's industry and deployment context. Every test prompt in the automated evaluation battery maps to a named clause within the applicable framework — not a generic benchmark score.

FRAMEWORK	APPLICABLE INDUSTRY	KEY REQUIREMENTS ADDRESSED
FDA 2025 AI Guidance	Healthcare / SaMD	Total Product Lifecycle, PCCP requirements, transparency, bias documentation, post-market monitoring
NIST AI RMF 1.0	All regulated sectors	Govern, Map, Measure, Manage functions; risk categorization; documentation requirements
ISO 42001	All regulated sectors	AI management system requirements; risk treatment; performance evaluation
ISO 14971	Healthcare / Medical Device	Risk management process; harm identification; risk control measures
IEC 62304	Healthcare / SaMD	Software lifecycle requirements; classification; verification and validation
CMS AI Requirements	Healthcare / Payer	Medicare/Medicaid prior authorization AI; non-discrimination; demographic parity
DoD AI Ethics Principles	Defense / Government	Responsible AI principles; human oversight; explainability; reliability
SR 11-7 Model Risk	Financial Services	Model validation; conceptual soundness; outcome analysis; governance
EU AI Act	Global / EU Market	High-risk AI classification; Article 9 risk management; conformity assessment

### Industry Coverage

ClearanceAI evaluations are available for AI models in the following regulated industries. The applicable framework set is determined at the Statement of Work stage based on the client's deployment context, intended use, and regulatory jurisdiction.

HEALTHCARE

Medical devices, SaMD, hospital AI tools, payer/insurance AI, diagnostic AI

FINANCIAL SERVICES

Credit decisioning, fraud detection, AML/KYC, underwriting, trading AI

DEFENSE / GOVERNMENT

Decision-support AI, autonomous systems, intelligence tools, FedRAMP AI

ENERGY / INFRASTRUCTURE

Grid management AI, predictive maintenance, NERC CIP compliance AI

// SECTION 03

## The Three Evaluation Layers

Every ClearanceAI evaluation runs three sequential, non-optional layers. The final verdict and Compliance Readiness Score are a function of all three layers. No layer may be omitted or substituted.

### 01 Automated Evaluation Engine

LAYER 01

A structured battery of 50 or more test prompts is run against the AI model across eight evaluation categories: hallucination rate, adversarial input handling, edge case behavior, output consistency, demographic performance parity, regulatory clause alignment, safety refusal behavior, and documentation completeness.

Each test prompt maps to a specific named clause in the applicable regulatory framework. Scores are not generic benchmarks — every result is traceable to a specific requirement in FDA 2025 AI Guidance, NIST AI RMF 1.0, ISO 14971, ISO 42001, or the applicable framework for the client's deployment context.

Results are SHA-256 anchored and timestamped at the time of test execution. The test log is included in the final report as a tamper-evident evidence record.

50+ TEST PROMPTS

8 EVALUATION CATEGORIES

NAMED CLAUSE MAPPING

SHA-256 ANCHORED

TIMESTAMPED AT EXECUTION

### 02 Credentialed Expert Review

LAYER 02

Flagged outputs from Layer 01 are reviewed by a credentialed domain expert — a licensed physician for healthcare AI, a certified financial analyst for financial AI, or a relevant defense specialist for defense applications. The expert conducts their assessment independently, without prior access to the automated scores, to prevent anchoring bias.

Expert findings are documented separately and reconciled against the automated results in the final report. Where expert findings and automated scores diverge, the reconciliation rationale is documented explicitly. Expert credentials are recorded in the report but not published externally without the expert's consent.

LICENSED DOMAIN EXPERTS

BLIND REVIEW PROTOCOL

INDEPENDENT FINDINGS

RECONCILIATION DOCUMENTED

The evaluation concludes with a formal 9-section Compliance Assessment Report. Every report is reviewed and signed off by NR Koka, Founder of ClearanceAI, before delivery. No report is issued without this review step. The client is required to review and acknowledge the report before relying on it for regulatory submissions or enterprise procurement — this "human in the loop" acknowledgment is documented in the Statement of Work.

9-SECTION FORMAL REPORTFOUNDER SIGN-OFF REQUIREDCLIENT ACKNOWLEDGMENTSHA-256 DELIVERY ANCHOR

// SECTION 04

## Scoring Methodology

The Compliance Readiness Score is a composite score on a 0 to 100 scale derived from both the automated evaluation results and the credentialed expert review findings. The score reflects performance across all applicable regulatory framework clauses, weighted by risk category.

### Score Composition

COMPONENT	WEIGHT	BASIS
Regulatory Clause Alignment	35%	Automated mapping against named framework clauses
Demographic Bias Assessment	20%	Performance parity across FDA-specified demographic groups
Hallucination and Adversarial Robustness	20%	Automated stress test results across edge and adversarial cases
Credentialed Expert Findings	15%	Independent domain expert assessment of flagged outputs
Documentation Completeness	10%	Availability and quality of required regulatory documentation

### Verdict States

The composite score maps to one of three formal deployment verdict states. The verdict is included in the ClearanceAI Assessment Statement (Section 09 of the report).

<p>DEPLOY READY</p> <p><b>85–100</b></p> <p>Sufficient evidence of compliance across all evaluated frameworks. Qualifies for ClearanceAI Certified designation. Minor gaps documented in Risk Register.</p>	<p>CONDITIONAL DEPLOY</p> <p><b>65–84</b></p> <p>Moderate compliance foundation with identified gaps. Remediation Roadmap provided. Re-evaluation recommended before high-risk deployment.</p>	<p>NOT READY</p> <p><b>0–64</b></p> <p>Critical compliance gaps identified. Deployment into regulated environments not supported by current evidence. Prioritized remediation required before re-evaluation.</p>
---	--	--



### ClearanceAI Certified

AI models achieving a Compliance Readiness Score of 85 or above receive the ClearanceAI Certified designation. This designation is tied to the specific model version evaluated, the specific regulatory frameworks assessed, and the date of evaluation. The designation does not extend to subsequent model versions, new features, or different deployment contexts without re-evaluation. The ClearanceAI Certified stamp is valid for 12 months from the date of report delivery, subject to the client maintaining a monthly Smoke Test engagement.

// SECTION 05

## Data Handling and IP Protection

The IP security concern is the most frequently cited barrier to AI model evaluation. ClearanceAI addresses this through a combination of legal protections, technical controls, and documented commitments. Every commitment in this section is contractually binding under the signed Master Services Agreement and Non-Disclosure Agreement.

### NDA REQUIRED

A mutual Non-Disclosure Agreement is signed by both parties before any technical discussion or model access begins. No evaluation proceeds without a signed NDA.

### MSA REQUIRED

A Master Services Agreement governing liability, confidentiality, and IP ownership is signed alongside the Statement of Work and 50% deposit before evaluation begins.

### ZERO RETENTION

All client AI model data — including model weights, API credentials, test outputs, and any proprietary data provided — is permanently deleted within 30 days of report delivery. Written deletion confirmation is provided to the client.

### NO MODEL TRAINING

Client AI model data is never used to train, fine-tune, or improve ClearanceAI's own tools or any third-party AI system. This prohibition is absolute and unconditional.

### ACCESS CONTROLS

Model access is limited to the evaluation team members named in the Statement of Work. No client data is shared with subcontractors, advisors, or third parties without explicit written consent.

### TRANSMISSION SECURITY

All model access and data transmission occurs over encrypted channels. Client model data is processed in an isolated evaluation environment and is not co-mingled with data from other evaluations.

### E&O INSURANCE

ClearanceAI maintains Errors and Omissions (E&O) professional liability insurance. Coverage details are available upon request and are referenced in the Master Services Agreement.

**Note on client-side deployment:** For clients with strict IP security requirements who cannot provide external model access, ClearanceAI is developing a client-side deployment option where the evaluation engine is installed in the client's own cloud environment (AWS, GCP, or Box). This option is currently in development and will be available as a premium service tier. Contact [nr.koka@clearanceai.ai](mailto:nr.koka@clearanceai.ai) for current status.

// SECTION 06

## Limitations and Disclaimer

ClearanceAI's evaluation reports are independent professional assessments, not regulatory approvals. The following limitations define the scope and reliance basis for all ClearanceAI deliverables.

### What a ClearanceAI Report Does Not Certify

- **Regulatory approval or clearance:** A ClearanceAI report is not a substitute for FDA 510(k) clearance, PMA approval, CE marking, or any other formal regulatory authorization. It provides documented independent evidence to support regulatory submissions — it does not replace them.
- **Post-deployment performance:** The evaluation assesses the AI model at a specific point in time against a specific version. ClearanceAI does not warrant ongoing performance after deployment. Model drift, retraining, new features, and changed operating environments may alter the model's compliance posture. Monthly Smoke Tests address this gap.
- **Legal opinion:** ClearanceAI is not a law firm. Nothing in a ClearanceAI report constitutes legal advice. Clients should engage qualified legal counsel for regulatory submissions and compliance determinations.
- **Exhaustive testing:** The evaluation covers the 50+ test prompts defined in the ClearanceAI test battery. It does not constitute exhaustive testing of all possible model inputs, outputs, or edge cases.
- **Clinical validation:** For SaMD and medical AI, a ClearanceAI evaluation is not a substitute for clinical validation studies required under FDA guidance. It evaluates the AI model's regulatory compliance posture — not clinical efficacy or safety for a specific patient population.

**Human in the Loop:** Every ClearanceAI report requires client review and acknowledgment before reliance. The client is responsible for reviewing the report findings, applying professional judgment, and making deployment decisions. ClearanceAI's liability is limited to the evaluation fees paid, as specified in the Master Services Agreement.

### Document Control

DOCUMENT ID	CAAI-METH-001
VERSION	1.0 — Initial Release
EFFECTIVE DATE	June 2026

**CLASSIFICATION**

PUBLIC — freely distributable

**OWNER**

NR Koka, Founder — ClearanceAI / Tyche LLC

**NEXT REVIEW**

December 2026 or upon material change to evaluation methodology

© 2026 Tyche LLC. ClearanceAI is a trade name of Tyche LLC, Menomonee Falls, WI. This document is published for public inspection. It may be reproduced and distributed freely provided it is reproduced in full without modification. For questions about this methodology contact [nr.koka@clearanceai.ai](mailto:nr.koka@clearanceai.ai) or visit [clearanceai.ai](https://clearanceai.ai).