

AI RISK AUDIT — SAMPLE OUTPUT

AI Risk Audit: Sample Output

Healthcare AI — Anonymized Demonstration

This is a redacted sample of a completed Ikwe Risk Audit applied to internal healthcare AI systems. It demonstrates the exact deliverables, methodology, and artifacts that clients receive.

System names, specific prompt data, and proprietary implementation details have been anonymized. All findings, scores, and remediation outcomes reflect real data.

Audit Date	February 2026	Sector	Healthcare AI
Systems	3 AI Systems (Anonymized)	Methodology	EQ Safety Benchmark v2.1
Data Basis	948 responses / 79 scenarios	Distribution	Redacted for External Use
Prepared By	Ikwe.ai	Classification	Sample — Not Confidential

ikwe.ai · stephanie@ikwe.ai · Visible Healing Inc., Iowa

EXECUTIVE SUMMARY

Audit Overview

This audit methodology surfaces risks that founders cannot reliably see from inside their own systems.

The audit surfaced three predictable failure classes across all three systems:

- **Authority drift** normalized by founder context — systems used clinical-adjacent language users interpreted as diagnosis
- **Emotional escalation** that founders could self-regulate but users could not — systems stayed present beyond clinical safety thresholds
- **Founder-as-safety-mechanism** — undocumented manual intervention acting as an invisible kill switch for three months

These are not edge cases. They are **structural failure modes** — and they are correctable.

RISK SCORECARD

Risk Assessment — Baseline

OVERALL RISK POSTURE

HIGH

Emotional escalation and dependency risks require structural mitigation before scaling

Risk Dimension	Severity	Likelihood	Scale Mult.	Score	Status
Emotional Escalation	5	4	4	80	HIGH
Dependency Formation	4	4	3	48	MEDIUM
Authority Drift	4	3	3	36	MEDIUM
Scale Amplification	4	3	4	48	MEDIUM
Governance Failure	3	4	2	24	LOW

Key Findings from 948 Evaluated Responses

54.7% of baseline responses introduced emotional risk despite appearing supportive

43% showed no repair behavior after causing harm

Models with higher emotional articulation performed worse on safety measures

REMEDIATION OUTCOMES

Measured Results — Before / After

After applying the Ikwe remediation framework — replacing founder intuition with system governance:

Risk Dimension	Baseline	Post-Mitigation	Reduction
Emotional Escalation	HIGH (80)	MEDIUM (40)	-50%
Dependency Formation	MEDIUM (48)	LOW (20)	-58%
Authority Drift	MEDIUM (36)	LOW (18)	-50%
Scale Amplification	MEDIUM (48)	MEDIUM (28)	-42%
Governance Failure	LOW (24)	LOW (8)	-67%

The fixes were structural, not intuitive — which means they transfer to any team.

This is one page from a complete, board-ready AI risk audit.

Request an audit for your AI systems.

Healthcare · Policy & Gov · LLM & Agents · Workplace & HR · Consumer

ikwe.ai · stephanie@ikwe.ai

About Ikwe.ai — Visible Healing Inc. builds behavioral emotional safety infrastructure for conversational AI. Recognition is not safety.