AI Risk Audit — Preview Pack

Healthcare AI — Anonymized Internal Deployment

This Preview Pack provides expanded visibility into a completed Ikwe Risk Audit — including the full risk scorecard, scoring methodology, and baseline assessment. System names are anonymized. Operational details are reserved for higher tiers.

Systems	3 AI Systems (Anonymized)	**Methodology**	EQ Safety Benchmark v2.1
Data Basis	948 responses / 79 scenarios	**Access Level**	Preview Pack

ikwe.ai/audit · stephanie@ikwe.ai

What the Audit Revealed

Founders are structurally blind. Not from incompetence — from proximity.

• **Authority drift** normalized by founder context — systems presented clinical-adjacent language users interpreted as diagnosis

• **Emotional escalation** that founders could self-regulate but users could not

• **Founder-as-safety-mechanism** — undocumented manual intervention acting as an invisible kill switch

54.7% introduced emotional risk	**43%** no repair after harm	**42–67%** risk reduction post-Ikwe

Baseline Risk Assessment

OVERALL RISK POSTURE

HIGH

Emotional escalation and dependency risks require structural mitigation before scaling

Risk Dimension	Severity (1–5)	Likelihood (1–5)	Scale Multiplier	Risk 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

Scoring: Risk Score = Severity × Likelihood × Scale Multiplier · 1–25 LOW · 26–62 MEDIUM · 63–125 HIGH

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

Baseline regulation: 1.7/5 → Post-mitigation: 4.05/5

Risk Reduction — Before / After

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 — they transfer to any team.

OBSERVED RISK EVENTS · ■ LOCKED

5 Documented Risk Events

RE-001 · Emotional Escalation — HIGH

Trigger, description, and evidence
Available in Playbook ($5,000)

RE-002 · Authority Drift — MEDIUM

Trigger, description, and evidence
Available in Playbook ($5,000)

RE-003 · Dependency Formation — MEDIUM

Trigger, description, and evidence
Available in Playbook ($5,000)

RE-004 · Governance Failure — LOW

Trigger, description, and evidence

Available in Playbook ($5,000)

RE-005 · Escalation Loop — MEDIUM

Trigger, description, and evidence

Available in Playbook ($5,000)

■ **Failure Mode Map — 4 Critical Failure Paths**

Available in Playbook ($5,000)

■ **Remediation Framework — Issue → Fix Table**

Available in Playbook ($5,000)

Unlock full risk events, failure modes, and remediation

Playbook: $5,000 · Full Audit: $25,000

ikwe.ai/audit