

AI Risk Audit

What we found when we audited our own AI — before auditing anyone else's. This preview shows high-level findings. Full details unlock at higher tiers.

WHAT YOU SEE HERE

- High-level findings and risk posture
- Before/after risk reduction summary
- Methodology overview
- Deliverable descriptions

WHAT UNLOCKS AT HIGHER TIERS

- **Full risk scorecard with scoring math**
- **Observed risk events with triggers & evidence**
- **Failure mode map with scale effects**
- **Remediation framework & governance rules**
- **Company-specific implementation**

Audit Date

February 2026

Sector

Healthcare AI

Systems

3 AI Systems

Methodology

EQ Safety Benchmark v2.1

Data Basis

948 responses / 79 scenarios

Access Level

Public Preview

01 ikwe.ai

What We Found

Founders are structurally blind. Not from incompetence — from proximity.
Your expertise fills gaps users will fall through.

Across three production AI systems and 948 evaluated responses, the audit surfaced three structural failure classes:

Authority drift normalized by founder context

Systems presented clinical-adjacent language users interpreted as diagnosis. The founder's domain of expertise masked the gap.

Emotional escalation founders could self-regulate but users could not

Systems remained present beyond safe thresholds. Founders intervened manually — users had no such circuit breaker.

Founder-as-safety-mechanism

Undocumented manual intervention acting as an invisible kill switch. Safety depended on one person's availability, not structural governance.

54.7%

of baseline responses
introduced emotional risk
despite appearing supportive

43%

showed no repair
behavior after
causing harm

Risk Reduction After Applying Ikwe

Dimension	Baseline	Post-Mitigation	Δ
● Emotional Escalation	HIGH 80	MOD 40	-50%
● Dependency Formation	MOD 48	LOW 20	-58%
● Authority Drift	MOD 36	LOW 18	-50%
● Scale Amplification	MOD 48	MOD 28	-42%
● Governance Failure	LOW 24	LOW 8	-67%

Faded = baseline. Solid = post-mitigation. Shorter = safer.



What You're Not Seeing

This preview shows high-level findings. The full audit contains layered content that unlocks at each tier:

1 Why We Audited Ourselves	This Preview
Failure classes and founder blind spots	
2 Systems Under Audit	\$2,500
System inventory, exposure ratings, scale stages	
3 Risk Scorecard — Baseline	\$2,500
Five-dimension scoring with formula and posture assessment	
4 Observed Risk Events	\$5,000
Triggers, evidence, risk scores for 5 identified events (RE-001 through RE-005)	
5 Remediation Framework	\$5,000
Issue → miss → fix table with governance rules	
6 Failure Mode Map	\$5,000
Critical failure paths and scale effects	
7 Priority Actions	\$25,000
Company-specific NOW / NEXT / LATER plan	
Company-Specific Implementation	\$25,000
Your systems, your scores, your report	

Sample of Redacted Content

REDACTED

Risk Scorecard — Full Six-Dimension Assessment
[Available in System Blueprint \(\\$2,500\) or higher](#)

REDACTED

Observed Risk Events RE-001 through RE-005
[Available in Playbook \(\\$5,000\) or higher](#)

REDACTED

Company-Specific Implementation & Governance Rules
[Available in Full Audit \(\\$25,000\)](#)

Access Tiers

	System Blueprint	Playbook \$5,000	Full Audit \$25,000	Implementation \$50K+
Findings Summary	█	█	█	█
Before/After Table	█	█	█	█
Risk Scorecard	█	█	█	█
Observed Risk Events	—	█	█	█
Failure Mode Map	—	█	█	█
Remediation Framework	—	█	█	█
Sector-Specific Mappings	—	█	█	█
YOUR Company Audit	—	—	█	█
Board-Ready Report	—	—	█	█
Embedded Governance	—	—	—	█
Ongoing Re-Audits	—	—	—	█

Ready to See More?

System Blueprint (\$2,500) · Playbook (\$5,000) · Full Audit (\$25,000)

ikwe.ai/audit

Methodology

All findings derived from the EQ Safety Benchmark v2.1 — a two-stage evaluation protocol that measures behavioral safety across six dimensions. Stage 1 (Safety Gate) determines whether a response introduces emotional risk at first contact. Stage 2 (Quality Dimensions) evaluates regulation, repair, and stabilization over sustained interactions.

This audit evaluated 948 responses across 79 vulnerability scenarios, testing multi-turn interactions where standard content safety benchmarks show no signal. The evaluation framework is the only assessment tool that measures behavioral patterns across turns, not just single-response content filtering.

The question is not whether your AI can recognize emotions. The question is whether your AI's behavioral pattern creates safety or risk over time — and whether you can prove it to a board.

Ikwe.ai · Visible Healing Inc. · EQ Safety Benchmark v2.1 · February 2026. This is a public preview. Full audit contents are available at tiered access levels. Contact stephanie@ikwe.ai for enterprise inquiries.