Sector Annex — Healthcare & Safety-Critical Systems

Interpretation of the Hybrid Human–Agent Operating Standard (Non-Normative)

Annex purpose

This annex interprets the Hybrid Human–Agent Operating Standard for healthcare and other safety-critical domains. It does not modify, override, or extend the Standard. Its purpose is to clarify how decision authority, autonomy, accountability, and governance apply where harm is asymmetric, reversibility is low, and authority is licensed.

1. Sector context: why healthcare is different

Healthcare decisions are characterized by asymmetric risk, licensed authority requirements, high audit and traceability expectations, mixed clinical and non-clinical decision domains, and strong public and regulatory scrutiny. AI adoption often fails not due to model performance, but due to unclear boundaries between support, recommendation, and authority.

2. Decision domains in healthcare

Healthcare systems span patient education and guidance, symptom triage and routing, alerting and monitoring, clinician decision support, care pathway recommendations, and operational/administrative decisions. Each domain must be explicitly classified before autonomy is allocated. No healthcare AI system should operate without a decision inventory.

3. Autonomy allocation patterns

Common defensible patterns under the Standard include A1–A2 for non-clinical education, navigation, scheduling, and administrative support; A2–A3 for clinician-facing support, prioritization, and alert generation; and A0–A1 for irreversible clinical decisions, treatment selection, and consent interpretation under ambiguity. Progression requires governance maturity, not confidence.

4. Accountability and clinical authority

AI systems do not hold clinical authority. Clinical accountability remains with licensed clinicians; organizations remain accountable for system design. Ownership does not imply continuous oversight. Human-in-the-loop designs that create hidden labor or approval theater increase risk rather than reduce it.

5. Governance and evidence emphasis

Healthcare governance must emphasize deterministic escalation behavior, clear clinical thresholds (not raw probabilities), end-to-end auditability, reconstruction of decision context after the fact, and explicit separation of clinical and non-clinical logic. Model accuracy alone is insufficient evidence.

6. Emotional and psychological risk

Patient-facing AI introduces additional risks including anthropomorphization, emotional reliance, boundary erosion between support and advice, and reinforcement of harmful beliefs. These risks must be treated as decision-governance risks, not UX issues. Emotionally influential interactions should default to lower autonomy.

7. Common healthcare failure modes

Ambiguous distinction between information and advice; over-reliance on disclaimers instead of governance; silent autonomy regression after adverse events; clinicians absorbing accountability without authority; and treating regulatory compliance as governance sufficiency.

8. What not to automate (yet)

Until governance maturity is proven: final diagnosis, treatment selection, consent interpretation under ambiguity, end-of-life decision support, and mental-health-critical judgement.

9. What success looks like under the Standard

Healthcare systems aligned with the Standard make decision boundaries explicit, preserve clinical authority without overload, scale AI support without hidden labor, handle rare failures without panic, and improve outcomes while maintaining trust.

Closing note

Healthcare does not require less AI. It requires clearer authority, stronger governance, and better failure handling. The Hybrid Human–Agent Operating Standard exists to provide that clarity.