

Clinical AI Risk Classification Standard

Artifact 2 | NorthStar Health Network

NorthStar Health Network is a fictional regional health system comprising 12 hospitals, 70 outpatient facilities, approximately 18,000 employees, and roughly 3 million annual patient encounters. All systems, metrics, incidents, vendors, and findings in this portfolio are simulated.

Risk Classification Objective

Apply a consistent, documented method to determine whether each AI system is Low, Moderate, High, Critical, or Prohibited. Classification drives required controls and approval authority.

Scoring Factors

Factor	1 - Low	2 - Moderate	3 - High	4 - Critical
Patient / clinical impact	No patient effect	Operational effect	May influence care	Could materially affect diagnosis, treatment, or safety
Decision influence	Drafting only	Recommendation	Strong workflow influence	Direct or near-direct clinical action
Data sensitivity	No PHI	Limited internal data	PHI	Highly sensitive / broad longitudinal PHI
Failure detectability	Easy to detect	Usually detectable	May be missed	Failure may be silent or time-sensitive
Explainability need	Low	Moderate	High	Essential for clinical challenge / patient safety
Scale	Small internal use	Department	Multi-site	System-wide / high-volume clinical use
Human recourse	Immediate	Available	Variable	Limited or time-critical

Tier Rules

Tier	Example	Required governance response
Prohibited	Unvalidated autonomous diagnosis or treatment action; AI that bypasses required clinician review	Do not deploy; disable; escalate to Clinical AI Governance Committee
Critical	Sepsis prediction or other time-sensitive clinical decision support with material patient-safety implications	Formal impact assessment; independent clinical/technical validation; privacy/security review; go/no-go committee approval; enhanced monitoring; incident drill
High	Radiology prioritization, clinical summarization, patient-facing chatbot touching clinical issues	Impact assessment; specialty owner; human-oversight controls; targeted validation; monitoring and complaint review
Moderate	Bed forecasting, revenue-cycle AI	Owner approval; privacy/security review; defined controls; periodic monitoring
Low	Internal drafting assistant with no PHI or patient-care use	Approved tool list; acceptable-use rules; training; baseline monitoring

Escalation triggers

- Unexpected patient harm or near miss.
- Material performance decline or data drift.
- Subgroup outcome disparity that may affect safety or access.
- Unapproved vendor/model change.

- Use outside approved clinical context.
- Privacy/security incident involving PHI.
- Clinician complaints indicating automation bias, alert fatigue, or misleading output.

Reference basis

Structured around the NIST AI RMF Govern-Map-Measure-Manage model and healthcare-specific transparency, safety, privacy, and human-oversight concerns. Regulatory applicability must be evaluated per system function and deployment context.