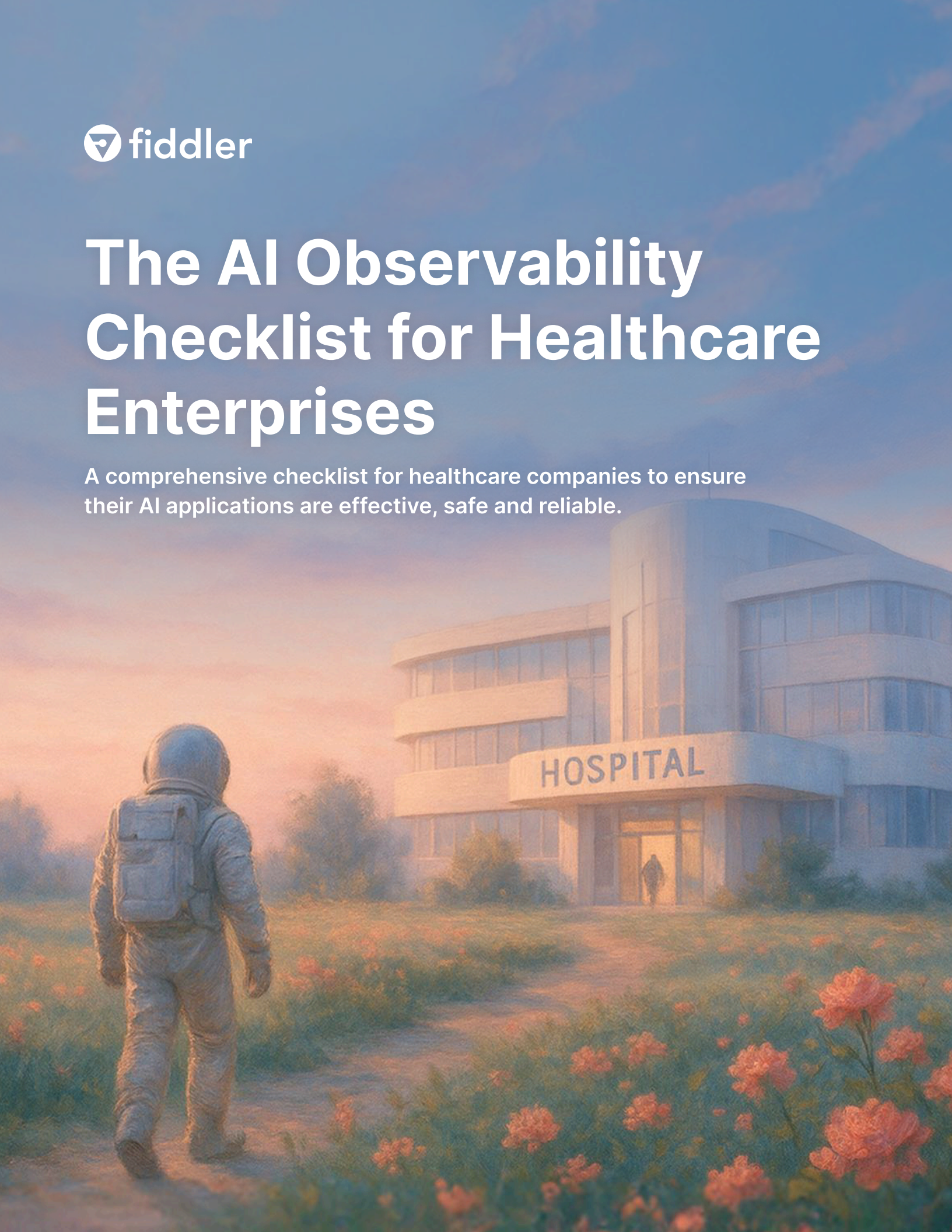
fiddler

The AI Observability Checklist for Healthcare Enterprises

A comprehensive checklist for healthcare companies to ensure their AI applications are effective, safe and reliable.



Monitoring AI in healthcare is a complex challenge with high stakes for patient safety, regulatory compliance, and brand trust.

This comprehensive checklist provides a roadmap for healthcare organizations to ensure their AI applications are effective, safe, and reliable.

Part 1: Foundational Setup & Risk Assessment

Establish a robust framework for AI oversight, risk management, and compliance to ensure alignment with your organization's clinical, ethical, and legal standards.

1. Establish a Cross-Functional AI Governance Committee

Appoint a multidisciplinary team (including clinical, data science, IT, legal, compliance, and patient safety) to oversee all AI systems. Ensure this committee has access to customizable dashboards that provide views into model performance, risk, and compliance.

2. Define Core Principles for Trustworthy AI

Codify the ethical principles that will guide all AI development and deployment. These principles should be the foundation for all AI policies and procedures. Key principles include:

Fairness & Equity
A commitment to identifying and mitigating biases that could worsen health disparities.

Accountability & Transparency
A clear understanding that clinicians retain ultimate responsibility, supported by explainable and auditable AI systems.

Patient Centricity & Safety
An assertion that the primary measure of success is a positive impact on patient outcomes.

3. Centralize Your AI Inventory

Maintain a comprehensive, centralized registry of all AI models, applications, and agents to enable enterprise-wide risk management and prevent unmonitored, siloed tools.

4. Define and Assign Risk Tiers to AI Applications

Assign risk tiers to all AI applications based on their potential impact on patient safety, clinical outcomes, and compliance (HIPAA, FDA) to tailor monitoring rigor accordingly. A patient-facing diagnostic agent carries a higher risk than an internal administrative chatbot, and monitoring frequency and rigor should be adjusted accordingly.

5. Adopt a Unified, Model-Agnostic Observability Platform

Deploy a unified monitoring platform, like Fiddler, that can securely integrate with your diverse production environments and model types. A unified platform provides a "single pane of glass" view, which is critical for effective governance, accelerating root-cause analysis, and creating a continuous feedback loop to improve AI systems.

6. Establish an AI Incident Response Protocol

Document a clear protocol for AI-related incidents, defining roles, responsibilities, and processes for investigation, resolution, and post-mortem review. This protocol should be supported by an observability platform that enables root-cause analysis (RCA) to accelerate the investigation and resolution of incidents.

Part 2: Pre-Deployment Evaluation & Go-Live Checks

Before any new AI application is deployed in a clinical setting, it must pass a rigorous series of validation checks to ensure it is safe, effective, and reliable for your specific patient population.

A. Validation for Predictive Models (ML)

1. Verify Data Integrity and Quality

Ensure training and validation data is complete, accurate, and representative. Implement automated checks for data quality (e.g., freshness, volume, distribution).

- [ ] **2. Conduct Robust Fairness and Bias Audits**

  Proactively test for performance disparities across relevant demographic subgroups. Document all results to meet regulatory expectations and ensure models do not perpetuate or amplify existing health inequities.

- [ ] **3. Establish Clinically Meaningful Performance Baselines**

  Validate model performance against a holdout dataset, analyzing not just overall accuracy but also metrics like precision and recall across patient cohorts. Work with clinical stakeholders to define the minimum acceptable performance thresholds required for safe real-world use.

B. Evaluation for Generative AI and Agents

- [ ] **4. Validate Against High-Risk Scenarios & Vulnerabilities**

  For LLMs and AI agents, proactively test for vulnerabilities before deployment. This includes stress-testing for clinical "hallucinations", potential PHI/PII leaks, and susceptibility to prompt injection attacks that could generate harmful or biased outputs.

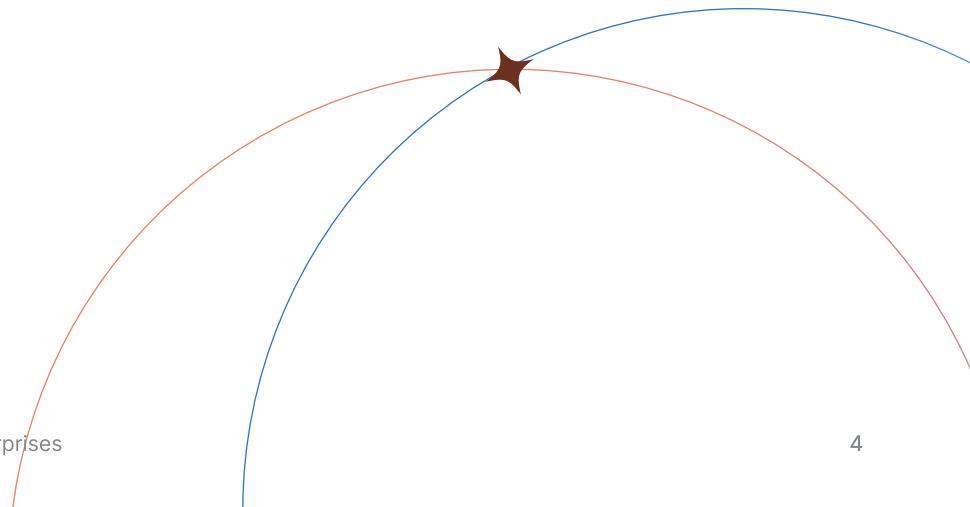
- [ ] **5. Establish Quality and Safety Baselines**

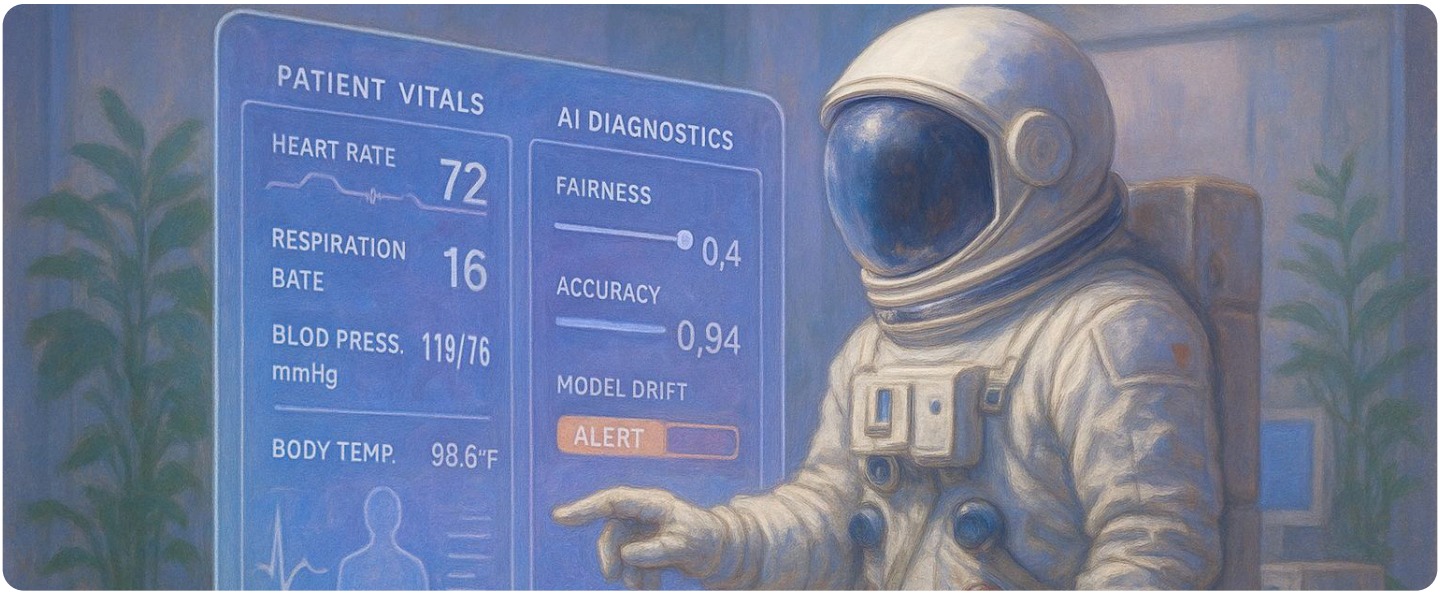
  Work with clinical and compliance stakeholders to define standards for acceptable output. This rubric should cover factual accuracy, clinical relevance, and safety, creating a clear baseline to validate against.

C. Final Governance Review

- [ ] **6. Conduct a Final Go-Live Review & Sign-Off**

  Require a final review and formal sign-off from the AI Governance Committee, confirming all pre-launch requirements for safety, fairness, and performance have been met.





Part 3: Continuous Production Monitoring & Governance

Once an AI application is live, monitoring is a continuous process. The goal is to maintain performance and trust by detecting issues in real-time, long before they can impact patient care.

A. Monitoring Predictive Models

- [ ] **1. Monitor for Performance Degradation**

  Configure automated alerts for any drops below clinically acceptable thresholds, and have tools in place to rapidly diagnose the root cause of the performance drop.

- [ ] **2. Implement Comprehensive Drift Detection**

  Monitor for statistical shifts in the model's environment, which serve as an early warning for future performance issues.

  Data Drift
  Track changes in input data distributions.

  Prediction Drift
  Monitor for changes in model output distributions.

  Concept Drift
  Detect when the relationship between inputs and outcomes changes.

B. Monitoring Generative AI and Agents

- [ ] **3. Activate Real-Time Guardrails for Generative AI**

  For LLMs and AI agents, implement automated guardrails to continuously monitor inputs and outputs. These systems should be configured to automatically block or flag harmful content, PHI/PII leaks, toxic language, and clinically significant hallucinations.

C. Ongoing Program Governance

- [ ] **4. Monitor Operational Health and Costs**

  Track infrastructure and service health metrics, including API latency, error rates, and throughput, to ensure the AI tool is responsive enough for clinical workflows. Correlate this with resource usage (CPU/GPU) and cloud costs to manage the financial sustainability of the AI program.

- [ ] **5. Institute a Quarterly AI Program Review**

  Have the AI Governance Committee conduct a regular review of the entire AI portfolio's performance, risk, and ROI. Use the unified observability platform as the single source of truth to ensure governance is an ongoing strategic function.