



Validation Strategies for Agentic AI Platforms in GxP-Regulated Clinical Data Environments

Presented by
Rajesh Hagalwadi

Maxis AI Inc, Edison, NJ

Session Objectives

Opening & Introduction

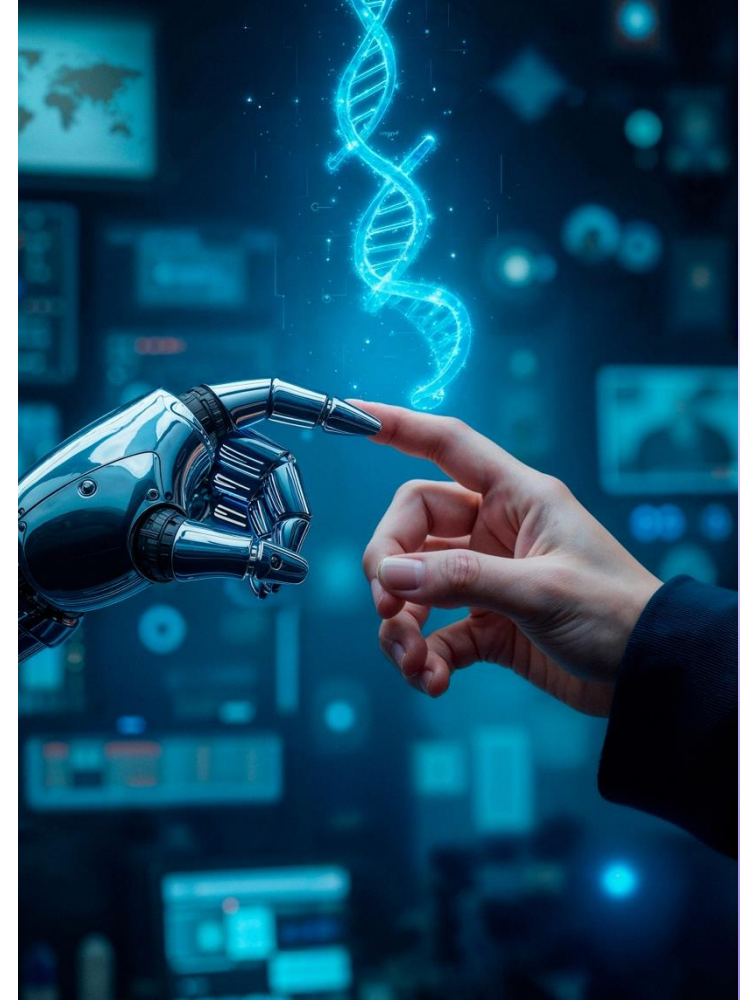
AI Hustle -Current
State & Adoption in
Life Sciences

What is Agentic AI

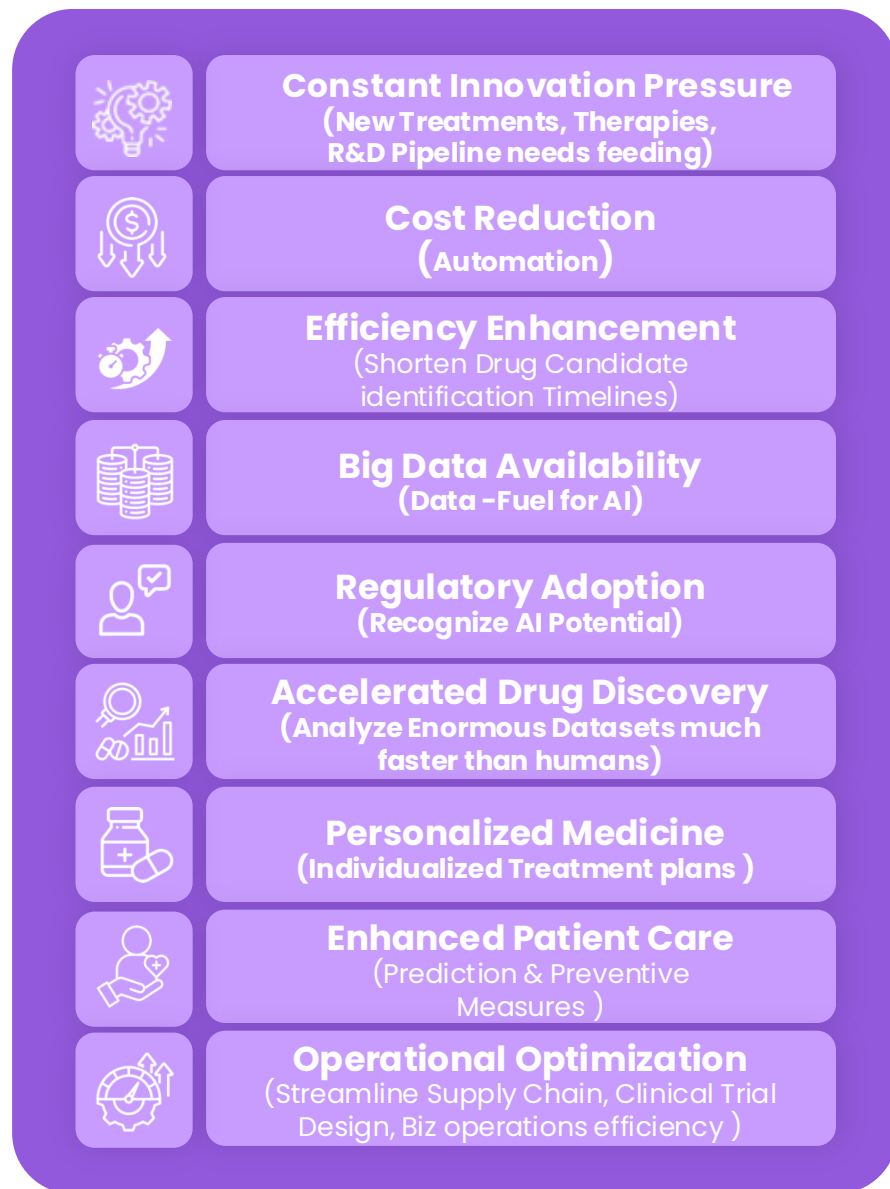
Agentic AI Validation Strategy

- AI Platform Validation Framework.
- AI Agent Lifecycle Phases for Validation, Safety, and Governance / Governance Framework for the AI agent Lifecycle.

An example to illustrate how it works in practice



Tailwinds (Driving AI Adoption Forward)



AI Adoption in Pharma & Biotech

Headwinds (Challenges)



**AI
Adoption**

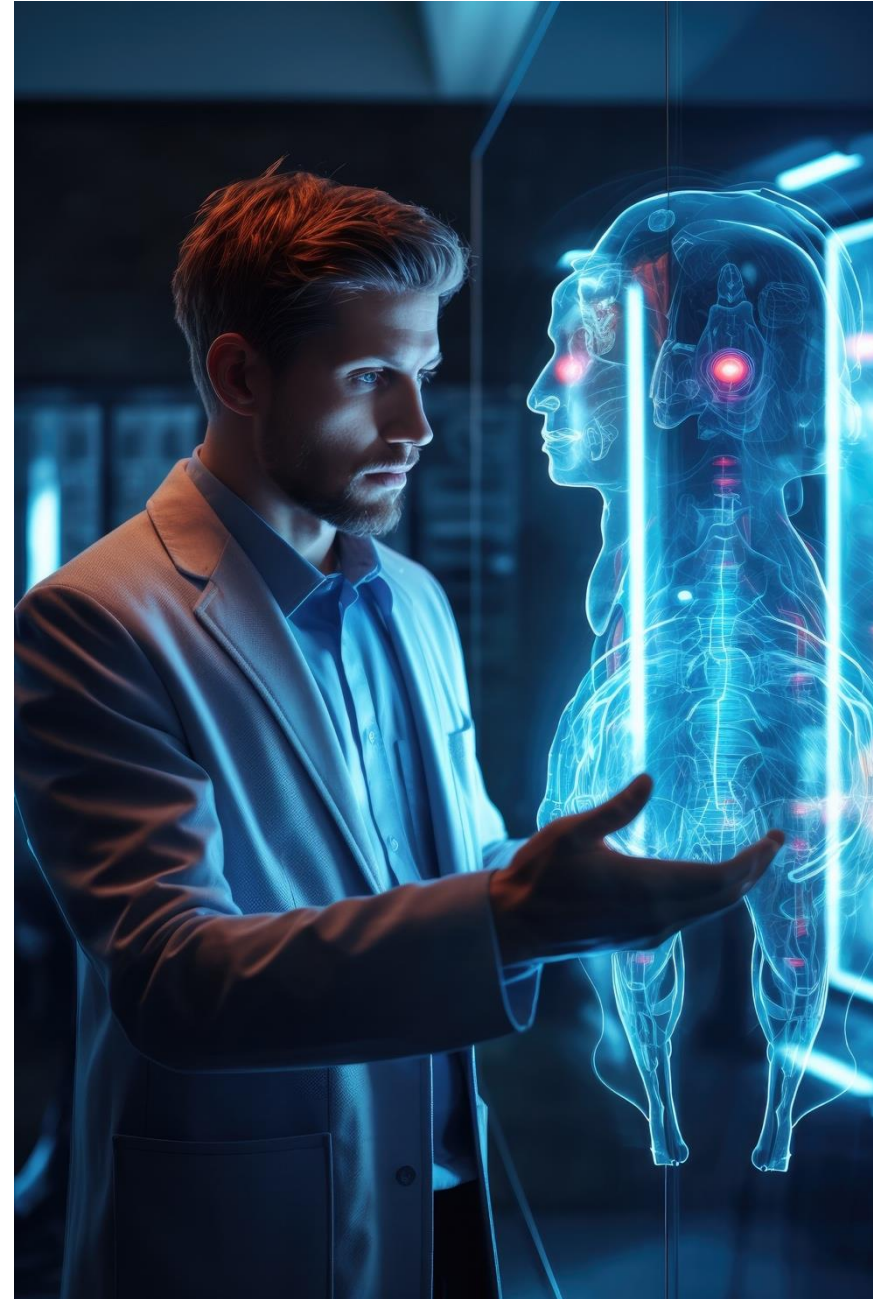
Strategic Outcome

Success in AI adoption depends on leveraging tailwinds while systematically addressing headwinds.

What is Agentic AI

Agentic AI is a coordinated system of AI agents, each with specialized capabilities, operating under an orchestration layer that plans, reasons and adapts workflows to achieve complex goals.

AI Agent: A system that can act autonomously by selecting a tool and retrieving data and memory to perform a task or achieve a goal.

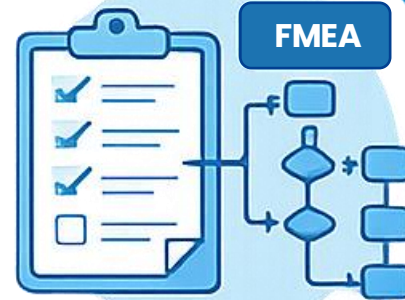


AI Platform Validation Framework



1. Concept Phase: Define & Assess

Document the System's intended use, conduct a high-level risk assessment and confirm regulatory applicability.

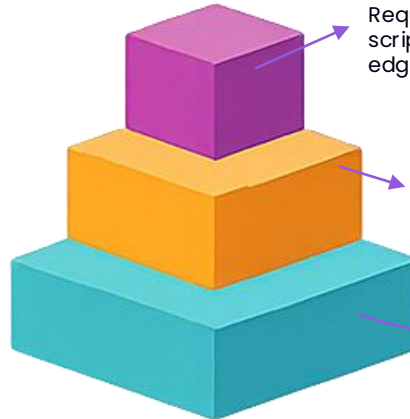


1. Concept Phase: Define & Assess

Document the System's intended use, conduct a high-level risk assessment and confirm regulatory applicability.

3. Development & Design Phase: Build & Test

Develop the platform and perform IQ and OQ Testing based on Risk Assessment.



Risk Priority Level: High

Required Testing Strategy: Robust, fully scripted testing covering all functional and edge-case scenarios.

Risk Priority Level: Medium

Required Testing Strategy: Partially scripted tests focused on key functionalities and known risk areas.

Risk Priority Level: Low

Required Testing Strategy: Unscripted or exploratory testing to validate usability and non-critical operations.



4. Acceptance & Release Phase: Verify & Deploy

Conduct final acceptance testing, formally release the platform and maintain full traceability documentation.

Platform Phase 1 & 2 : Concept and Requirements

01. Concept Phase (Definition)



Intended Use: Define functional scope, user base, and business impact.



Regulatory Check: 21 CFR Part 11 applicability.

02. Requirements Phase (Risk Stratification)

Mechanism: Failure Modes and Effects Analysis (FMEA).

High RP: Critical impact on patient safety/integrity.

Medium RP: Moderate impact on patient safety/integrity.

Low RP: Minimal impact on patient safety/integrity.

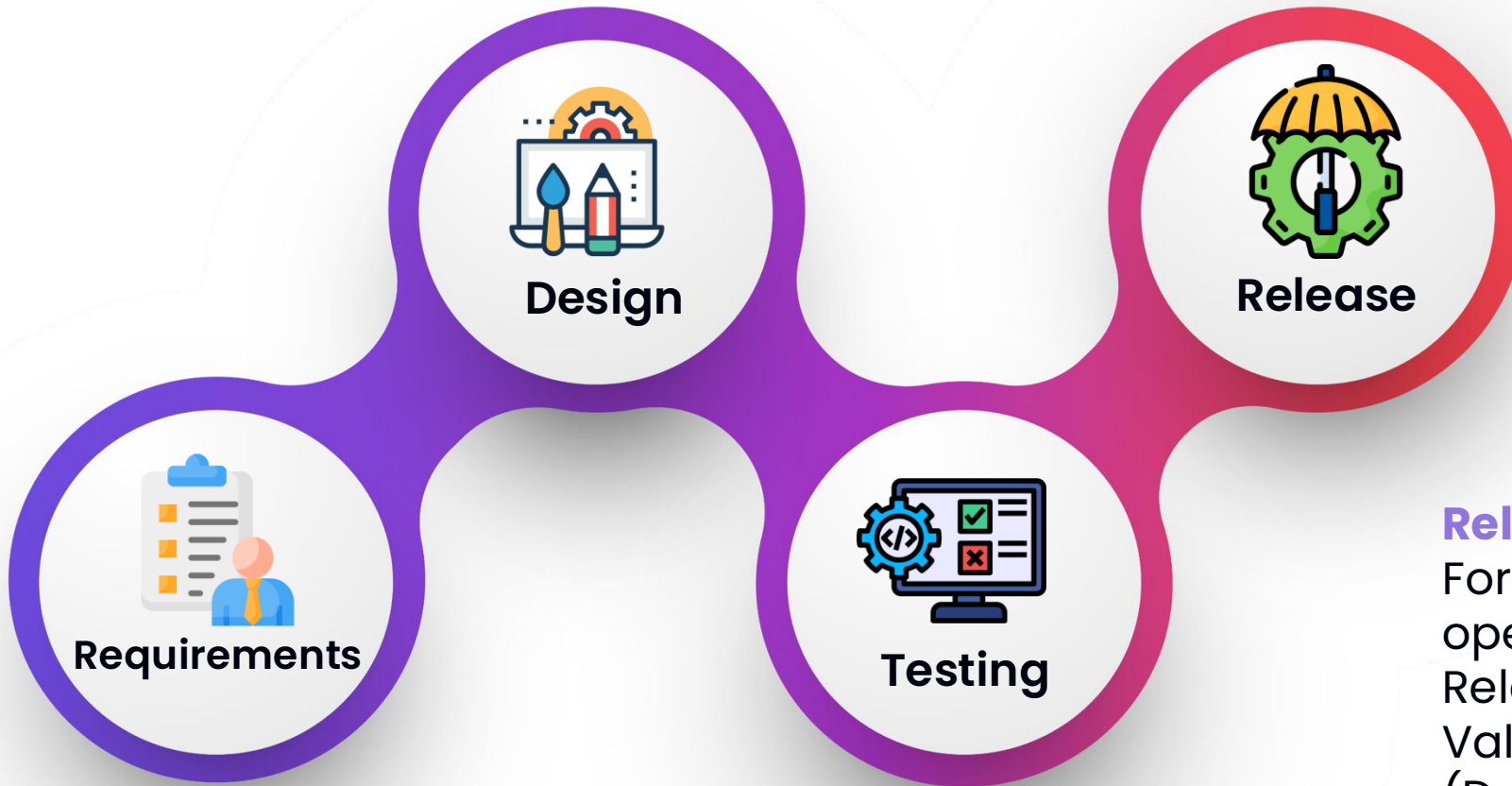
Platform Phase 3 : Dev , Design & Testing

Development Standard: Version control, secure data handling, documented architecture.

Testing Strategy (IQ/OQ)		
Priority Risk Classification	Strategy	Scope
High Risk Priority	Robust, fully scripted	All functional and edge-case testing.
Medium Risk Priority:	Partially scripted;	focused on key functionalities and known risks
Low Risk Priority	Unscripted / exploratory	Usability and non-critical functions

Key Insight: Testing effort is directly proportional to risk classification.

Platform Phase 4 : Acceptance and Release



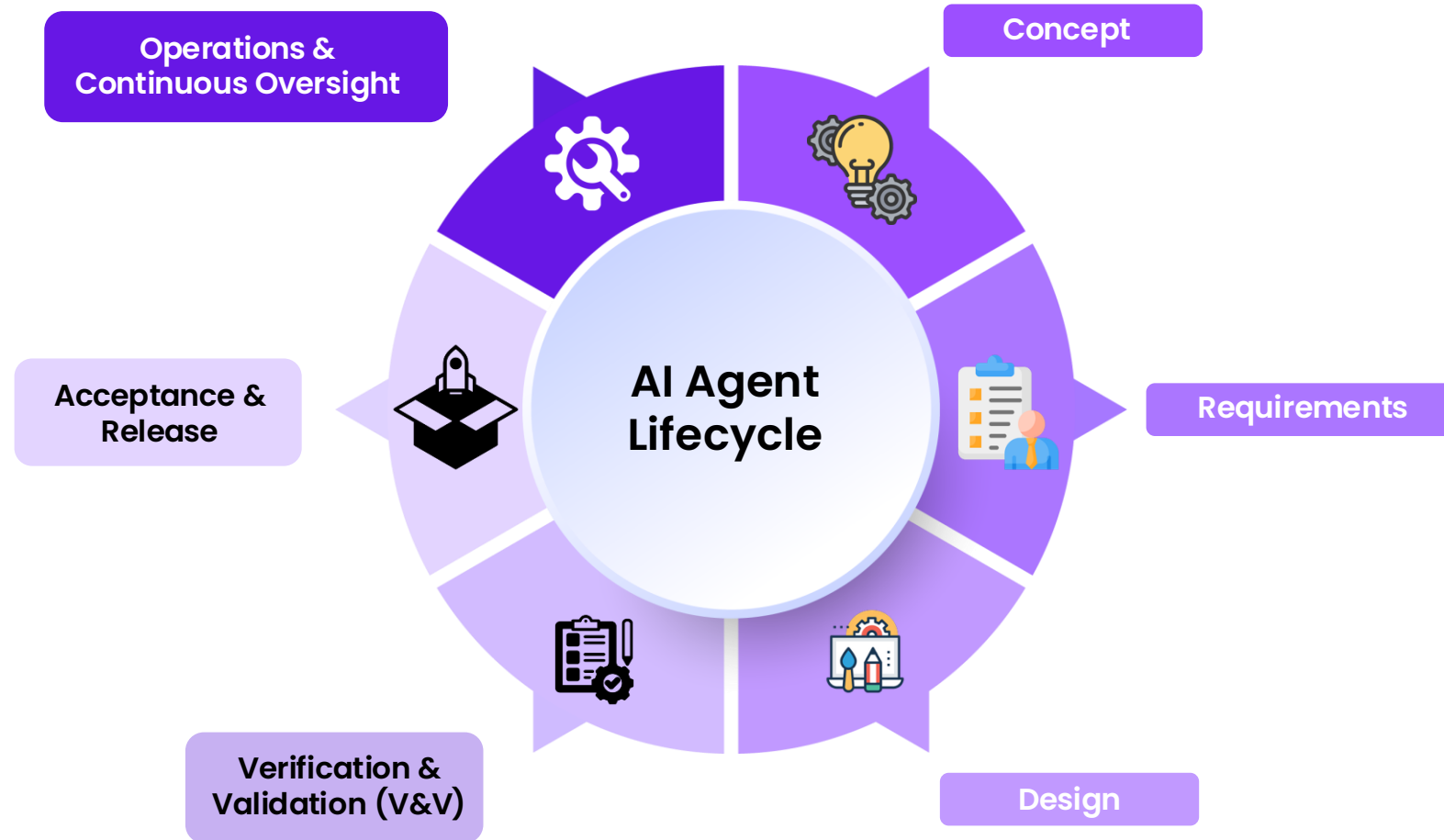
Traceability Chain:

Release Artifacts:

Formal approval for operational use
Release notes & configuration
Validation Summary Report
(Definitive Compliance Document)

Governance Framework for the AI Agent Lifecycle

End to End Strategies for Validation, Safety and GxP Compliance



Phase 1 : Concept – Establishing the Regulatory Foundation



1. **Intended Use Definition:**

Explicitly defining what the agent is for.

2. **Risk Assessment:**

Applying the Risk matrix to categorize (Model influence vs decision consequence) the agent's criticality.

3. **GxP Determination:**

Formal declaration of High, Medium, or Low GxP impact.

At this stage, we explicitly define intended use and risk to ensure development is anchored in ethical principles.

Concept

Requirements

Design

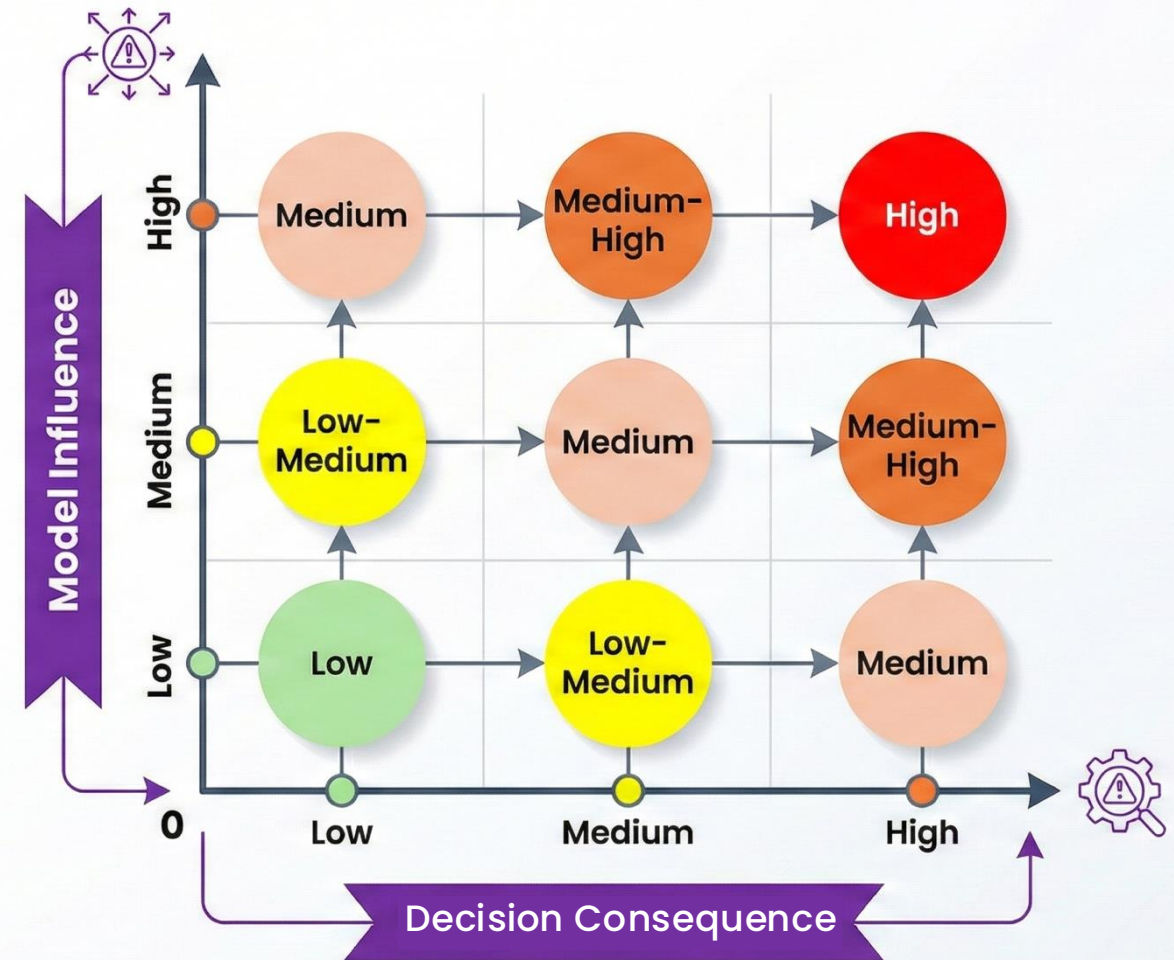
Verification and
Validation (V&V)

Acceptance
& Release

Operations & Continuous
Oversight

Agent Risk Assessment Matrix

Risk Matrix involves assessing model influence and decision consequence to define risk levels ranging from Low to High.



Phase 2 : Functional Requirements and Organizational Accountability

Traceability & Logging Specs

Traceability of an activity in the Clinical data Oversight & Monitoring process

```
{
  "timestamp": "2024-05-20T10:30:00Z",
  "log_level": "INFO",
  "patient_id": "PID12345",
  "vitals": {
    "BP": "120/80"
  },
  "query": "Evaluate risk of adverse event.",
  "reasoning_step": {
    "step_id": "R1",
    "model_version": "v1.2.3",
    "logic": "Analyzed vitals against protocol P-456."
  },
  "output": {
    "risk_score": "Low",
    "recommendation": "Continue standard monitoring."
  },
  "trace_id": "T-7890abcdef"
}
```

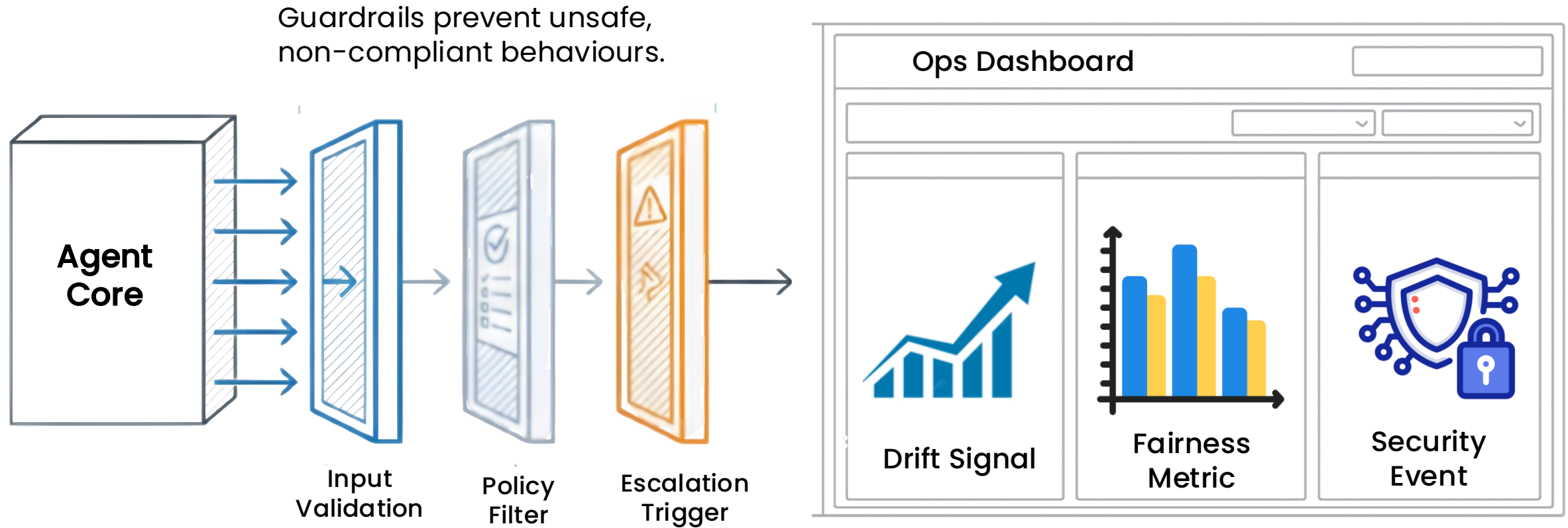
Comprehensive logging required to trace inputs, outputs, intermediate reasoning steps, and decision pathways. Enables post-hoc audits and incident investigations.

RACI Matrix

Role	R	A	C	I
Developer	✓			
QA Lead		✓	✓	
Product Owner		✓		✓

Ensures unambiguous responsibility for safety, monitoring, and legal compliance across technical, quality, and business stakeholders.

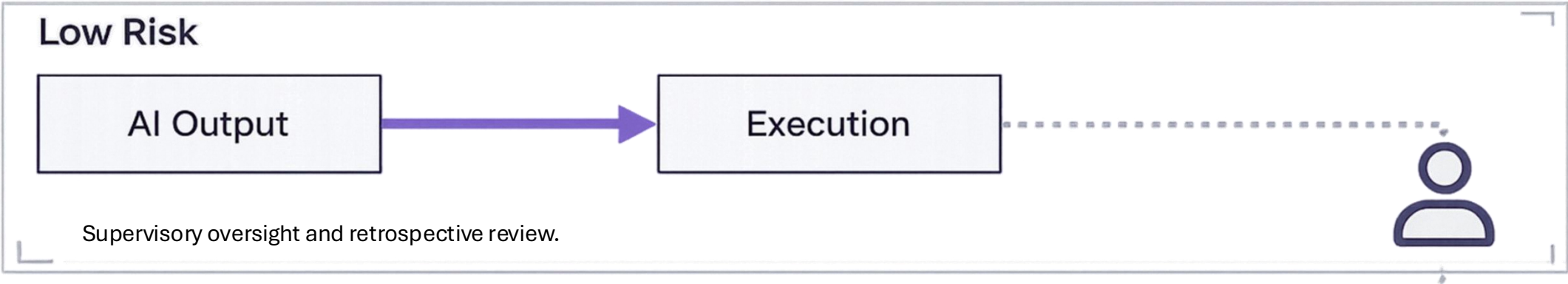
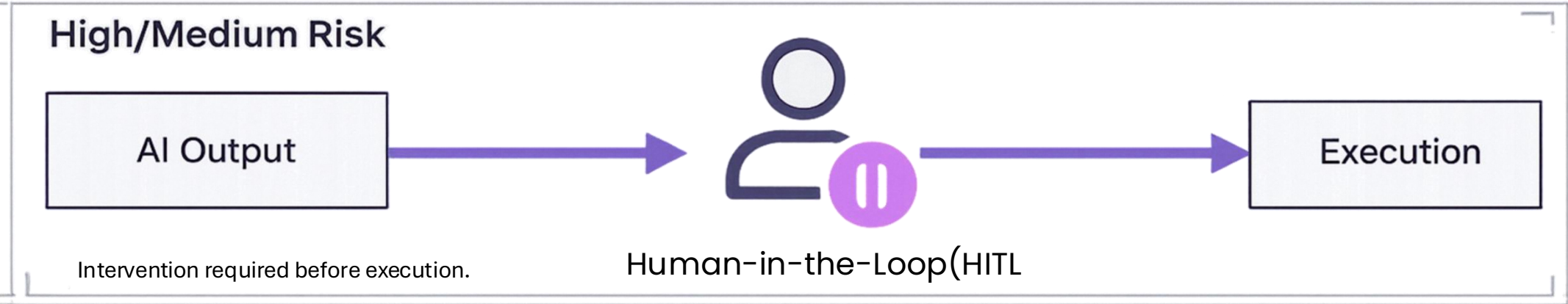




Architecture must favor interpretable models to document reasoning mechanisms.



Phase 3 (Deep Dive): Designing Proportional Human Oversight



Oversight mechanisms are designed proportional to the GxP impact defined in Phase 1

Human-on-the-Loop (HOTL)



Phase 4 : Verification & Validation- Generating Objective Evidence

Statistical Verification

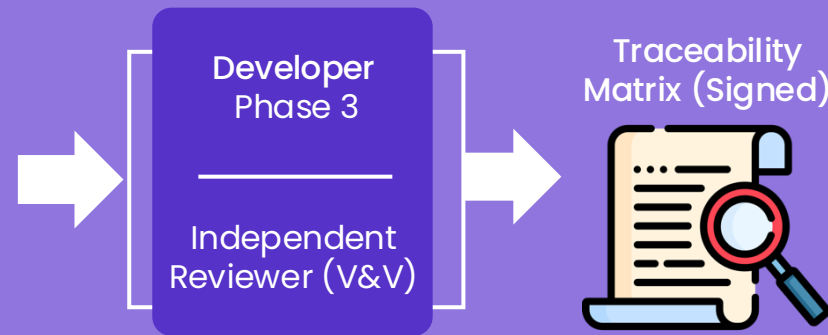
Use of accuracy, precision, recall, and confidence interval analysis.
Demonstrating consistency.

- Accuracy > 99.5%
- Precision = 98.2%
- Recall > 95.0%
- 95% Confidence Interval [0.94, 0.99]



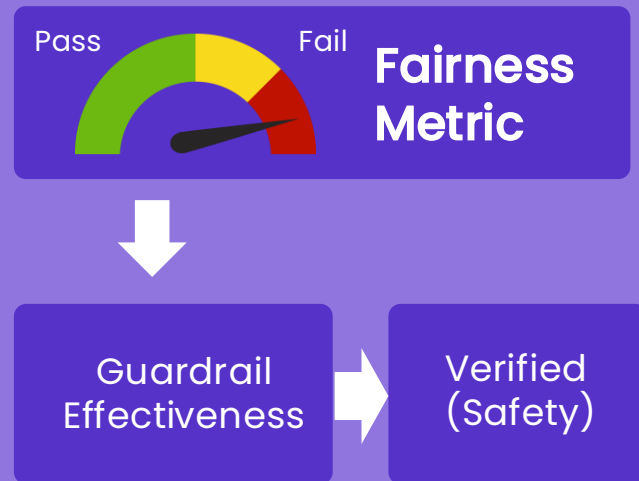
Independent Quality Review

Segregation of duties: The reviewer cannot be the developer.
Traceability matrices must be signed off.



Ethical Evaluation

Bias and fairness testing against Phase 2 thresholds. Guardrail effectiveness verification.



Concept

Requirements

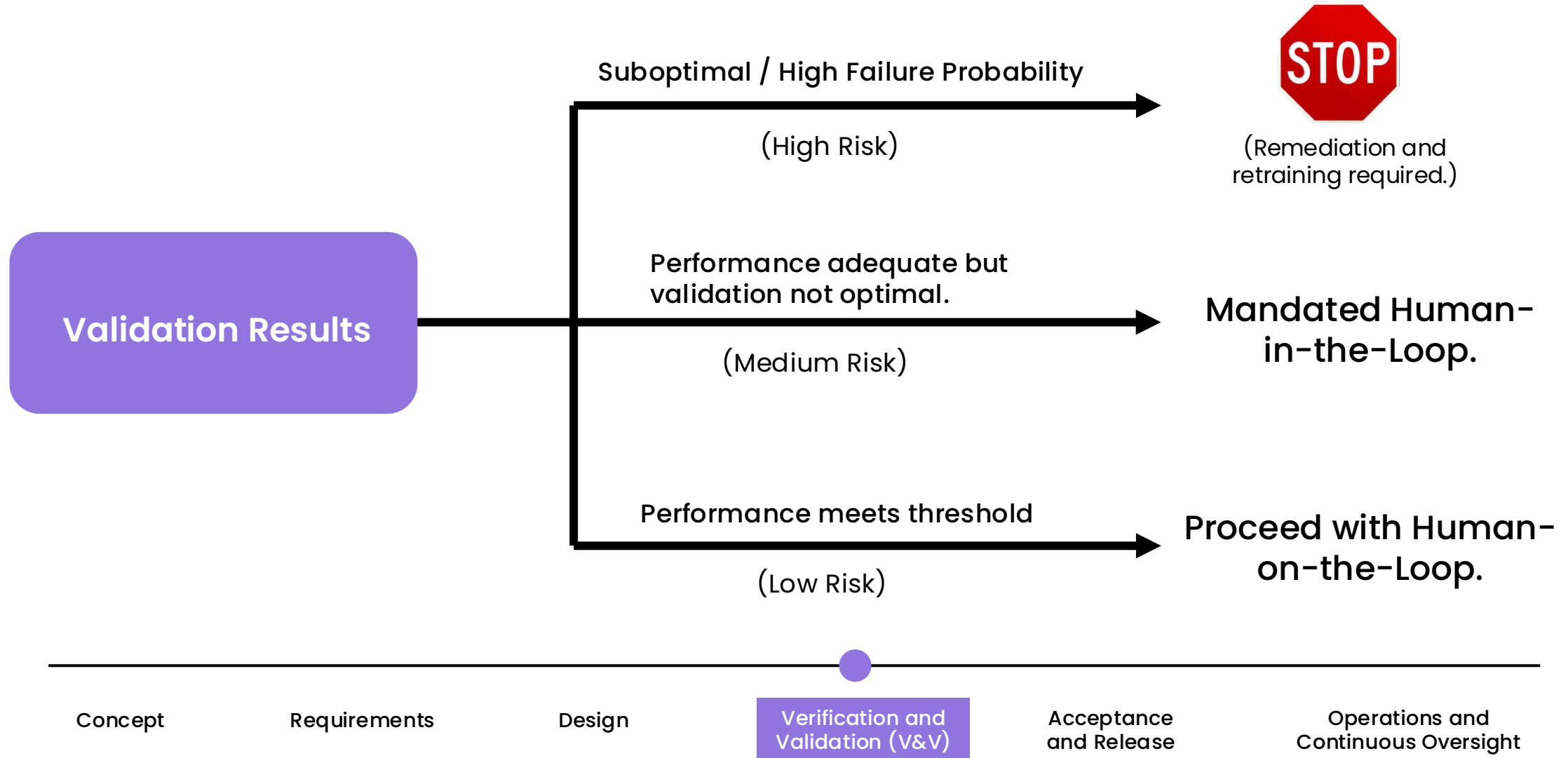
Design

Verification and
Validation (V&V)

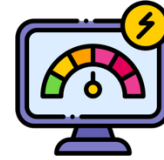
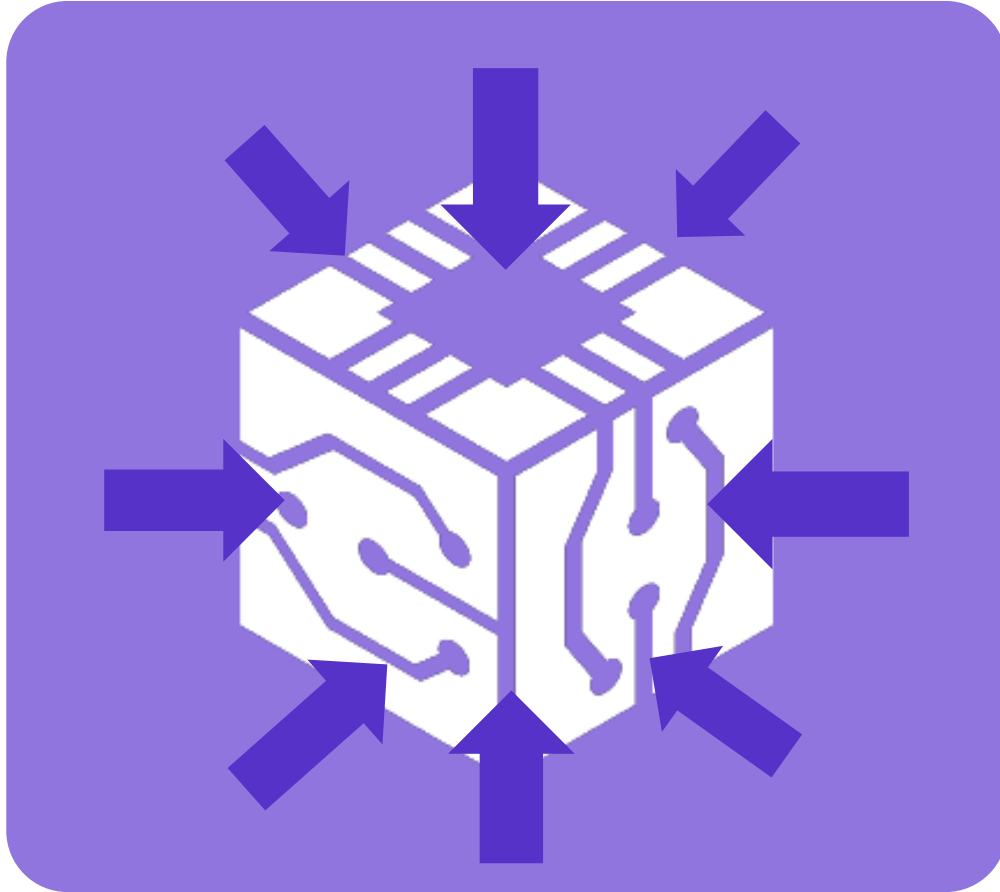
Acceptance
and Release

Operations and
Continuous Oversight

Risk-Based Performance Thresholds



MAXIS Stress Testing, Robustness, and Security



Stress Testing

Introduction of controlled failure conditions (noisy inputs, edge cases).



Adversarial Testing

Using adversarial prompts to assess resilience against manipulation.



Explainability Assessment

Verifying decisions can be justified to auditors.



Security Validation

Vulnerability scanning and access control checks.

Concept

Requirements

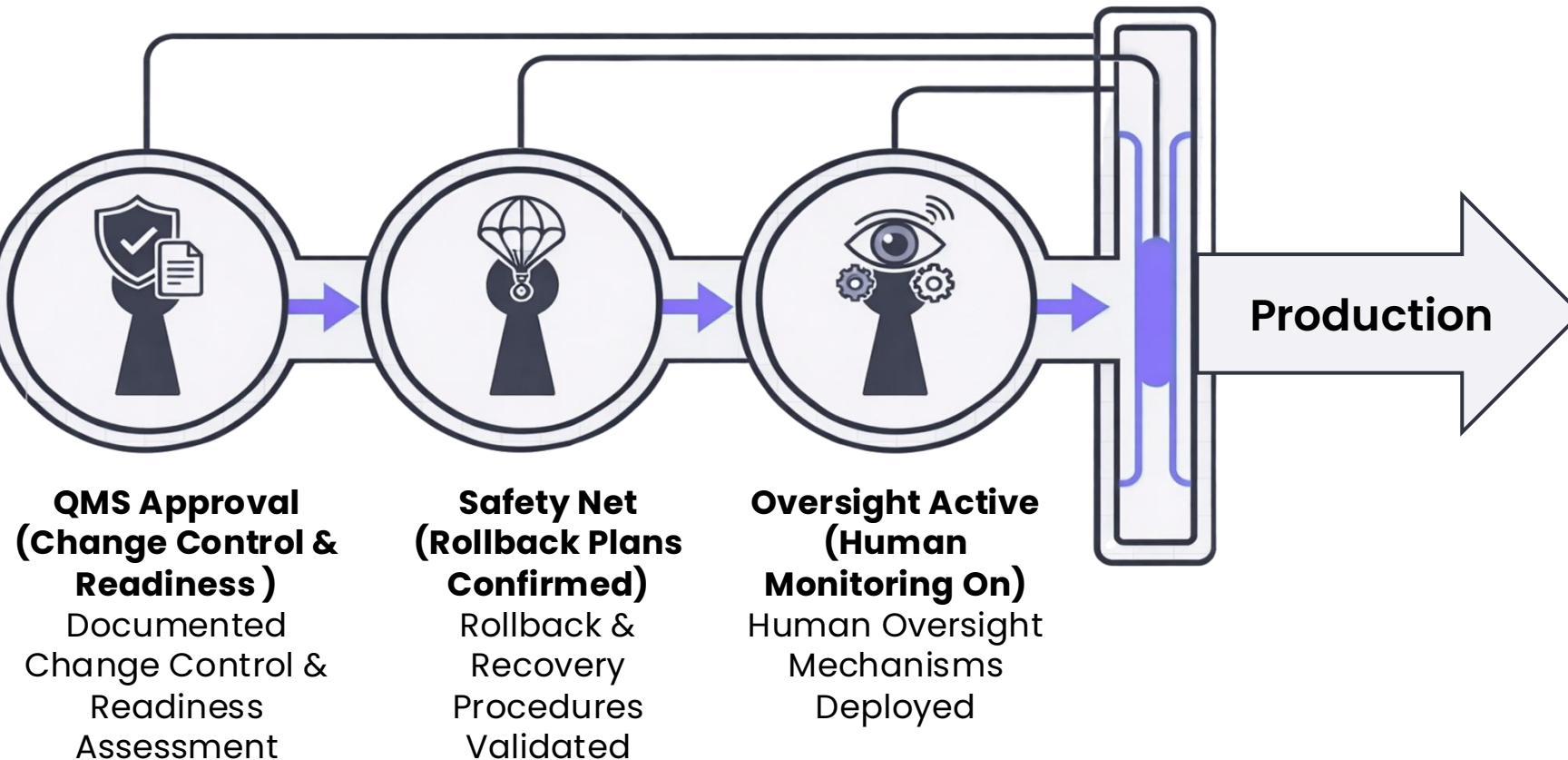
Design

Verification and
Validation (V&V)

Acceptance
and Release

Operations and
Continuous Oversight

Phase 5: Acceptance and Release — The Gate to Production



- Release is governed by structured Change Control.
- Requires documented approval, risk acceptance, and formal readiness assessment.
- Critical Oversight : Ensuring human oversight mechanisms are active for critical decision points immediately upon release.

Concept

Requirements

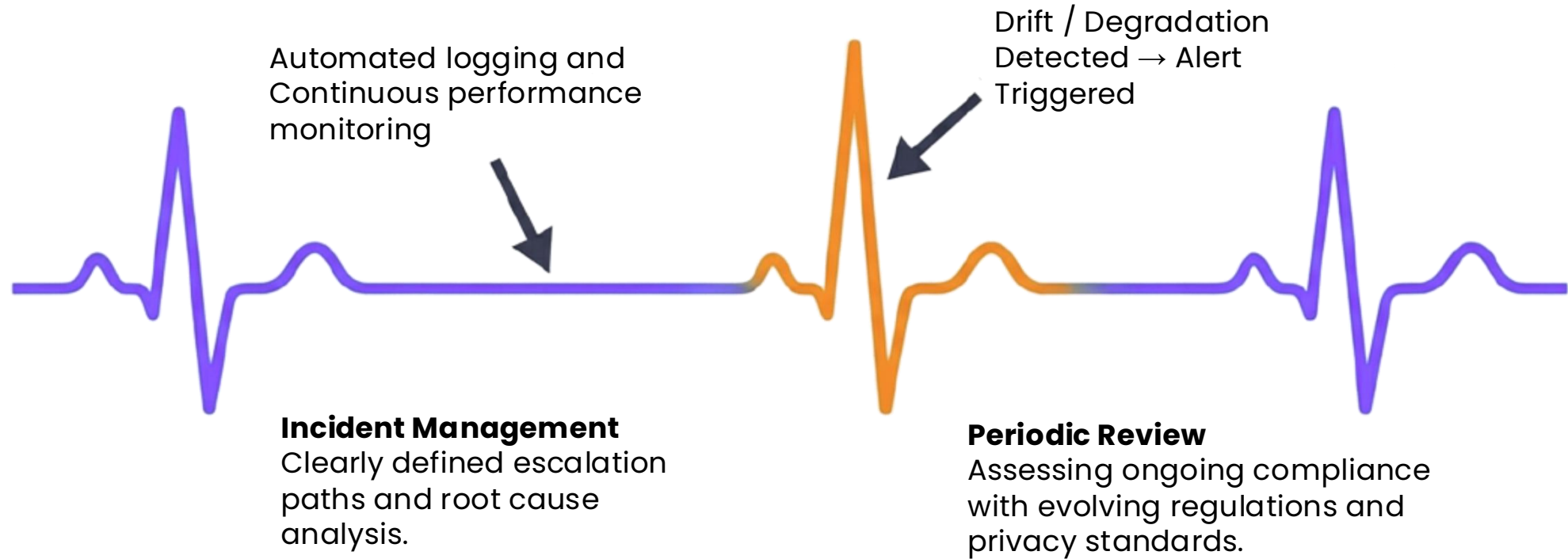
Design

Verification and
Validation (V&V)

Acceptance
and Release

Operations and
Continuous Oversight

Phase 6: **Operation** and Continuous Oversight



Concept

Requirements

Design

Verification and
Validation (V&V)

Acceptance
and Release

Operations and
Continuous Oversight

Lifecycle Summary Checklist

Concept	Requirements	Design	V&V	Ops
Intended Use — Defined	RACI — Established	Guardrails — Active	Statistical Validation — Pass	Drift Monitor — Active
Risk Matrix — Applied	Logging Specs — Defined	Model Oversight — Selected	Independent Review — Signed	Rollback Plan — Defined

“**Governance** is not a blocker, but the enabler of trust and long-term **system integrity**.”



Governance Mandate : Design for Compliance

**“Validation is not a retrospective activity .
It is the explicit definition of intended use,
risk and oversight before a single Line of
code is written.”**

**Traceability –
from
requirements to
reasoning**

**Accountability –
clear segregation
of duties**

**Continuity –
deployment is the
beginning of
surveillance**

Q&A

Thank you for your time!
Open to questions &
discussion.

Rajesh Hagalwadi



Maxis AI Inc.
510, Thornall Street,
Edison, NJ, 08837



Thornall Street



www.maxisai.com



732 494 2005



rajesh@maxisit.com