VALIDMIND®

AI Governance at Scale

David Asermely – VP of Global Business Development
Kristof Horompoly – Head of AI

VALIDMIND®

Modernize Governance and Scale Expertise

- Award-winning AI governance platform
- Enhances operational efficiency and drives long-term growth
- Promotes trust and transparency, strengthens regulatory compliance, and limits financial and reputational risks
- Fosters accountability and responsibility, increases customer satisfaction
- Global expertise & 24/7 customer support

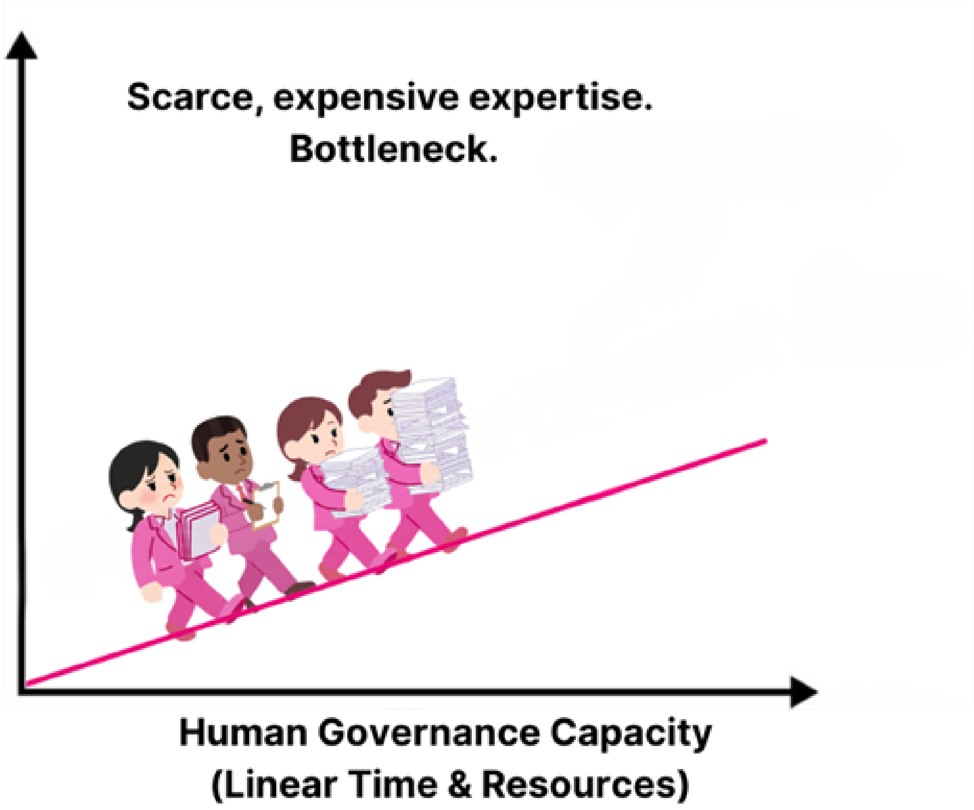
Chartis
RiskTech100
2026

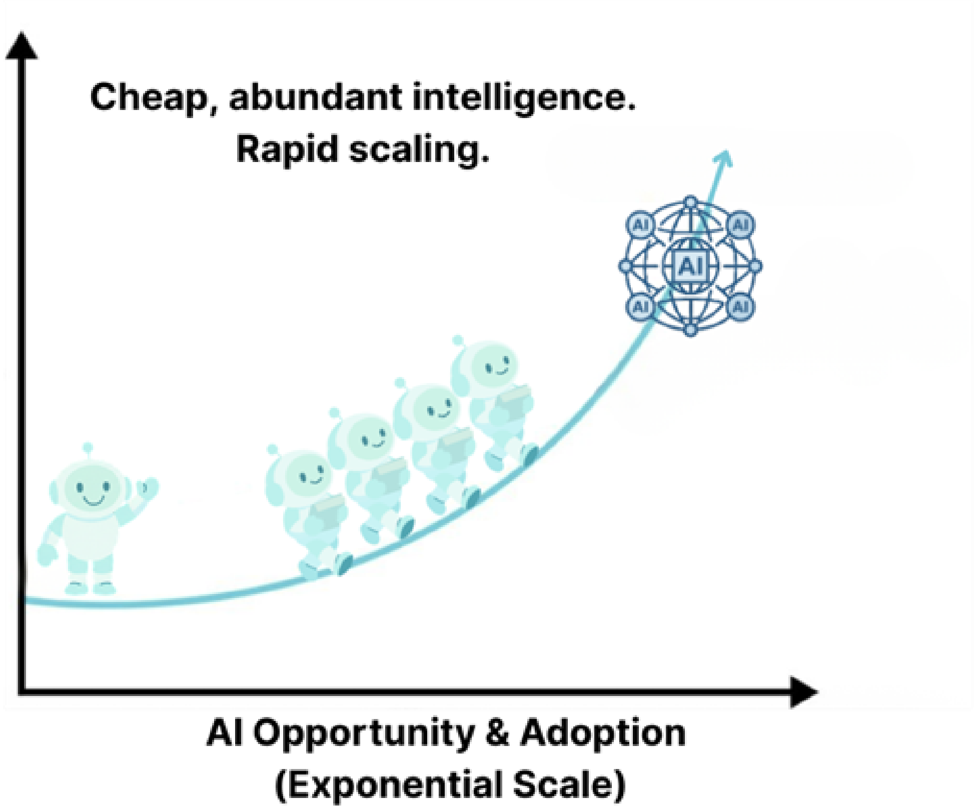
ValidMind
Artifical Intelligence
Governance

✓ Category win: Artificial Intelligence Governance

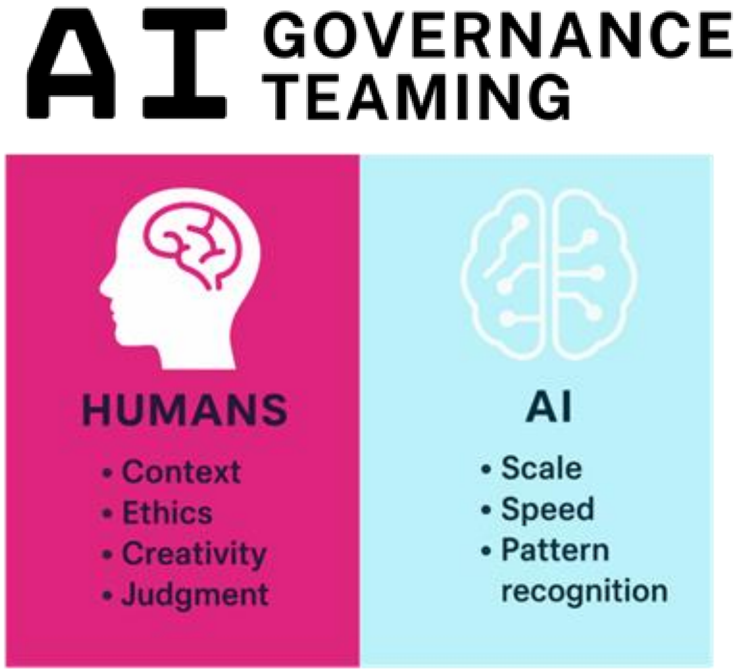
✓ Category win: Model Validation Supporting Tools

The Risk Gap: Exponential AI vs Linear Governance


Scarce, expensive expertise.
Bottleneck.
Human Governance Capacity
(Linear Time & Resources)



Enable AI Adoption with Governance Teaming



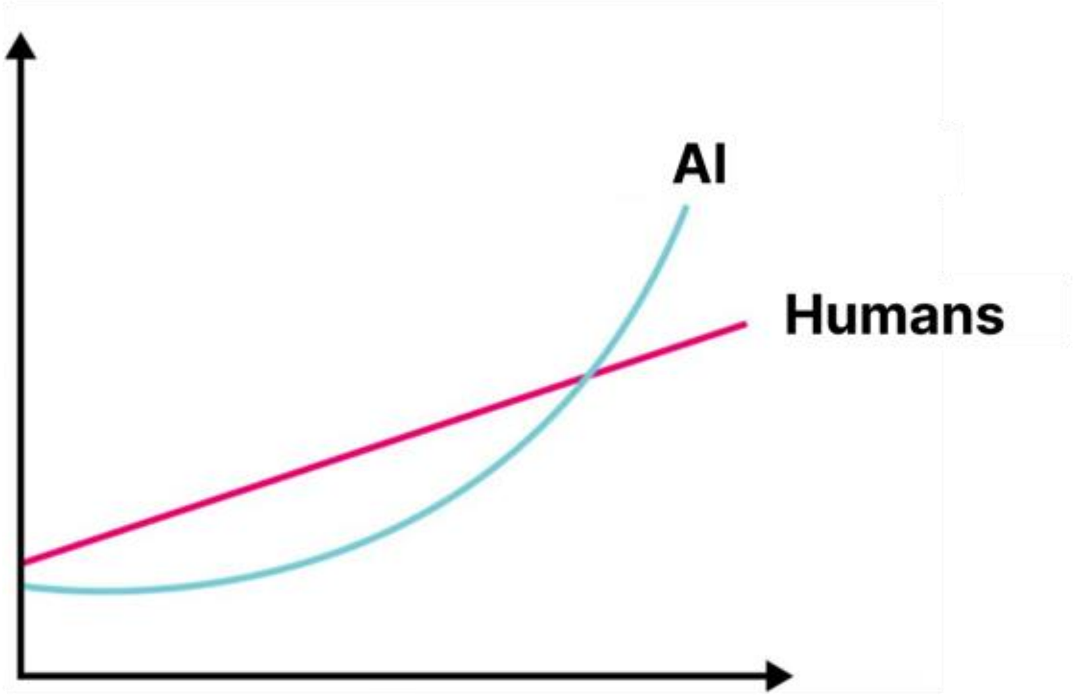
AI GOVERNANCE TEAMING

HUMANS

- Context
- Ethics
- Creativity
- Judgment

AI

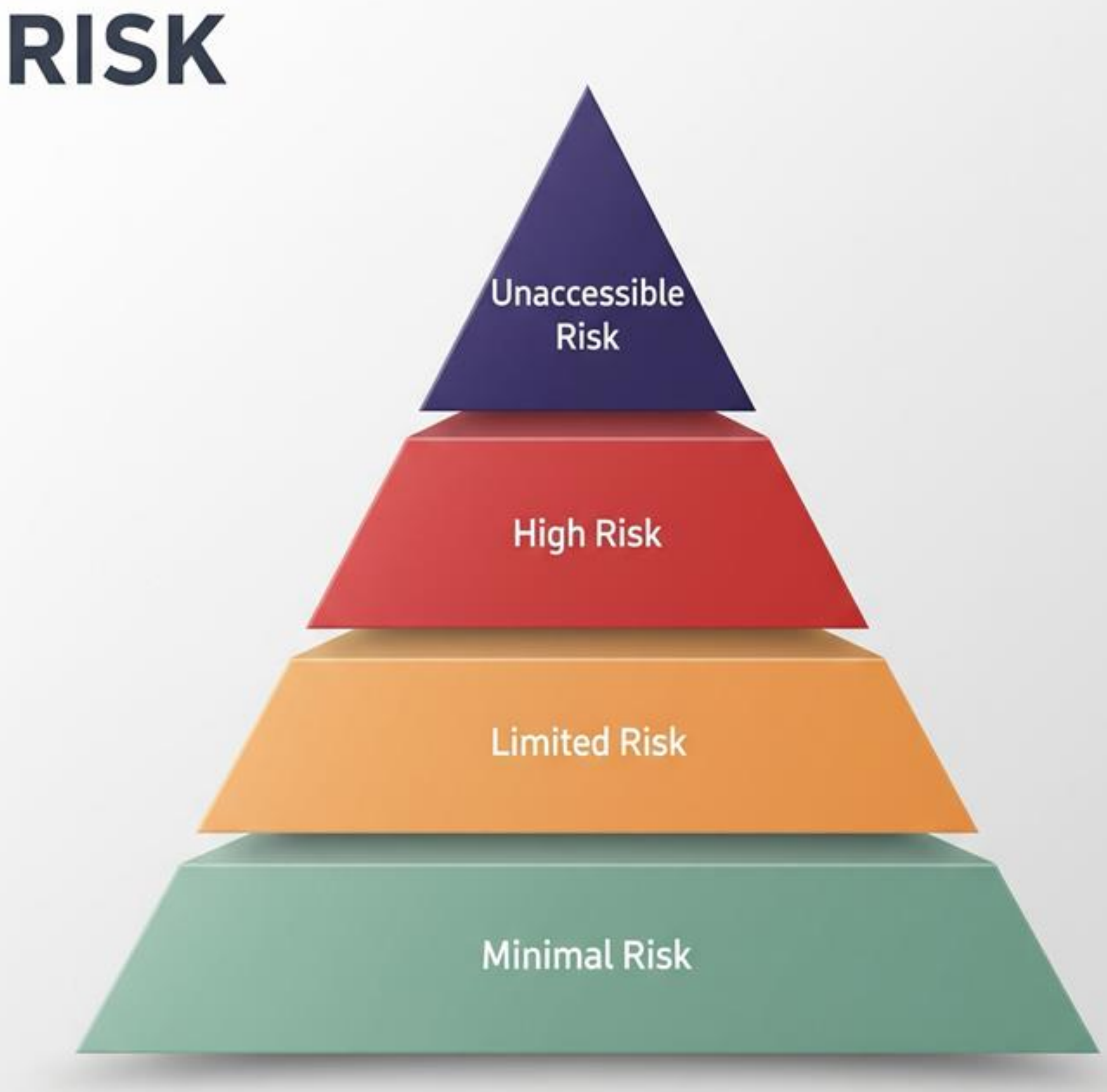
- Scale
- Speed
- Pattern recognition



SS1/23 Principle 1: Model Identification and Risk Classification

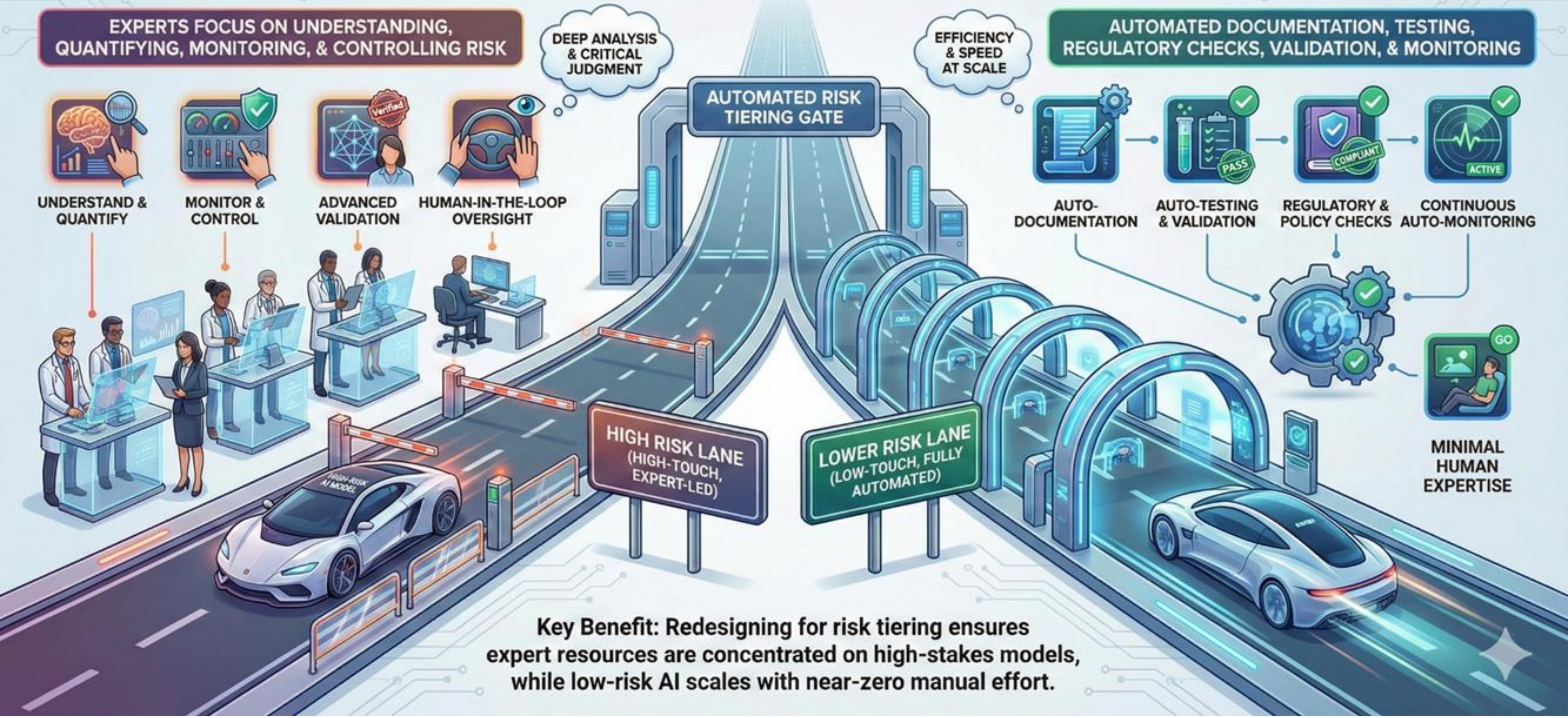
Humans can focus 80% of their time on high-risk models...

...and spend 20% of their time on low/medium-risk models, while our **automation handles the rest.**



THE REDESIGNED APPROACH: CONSISTENT & REPEATABLE RISK TIERING

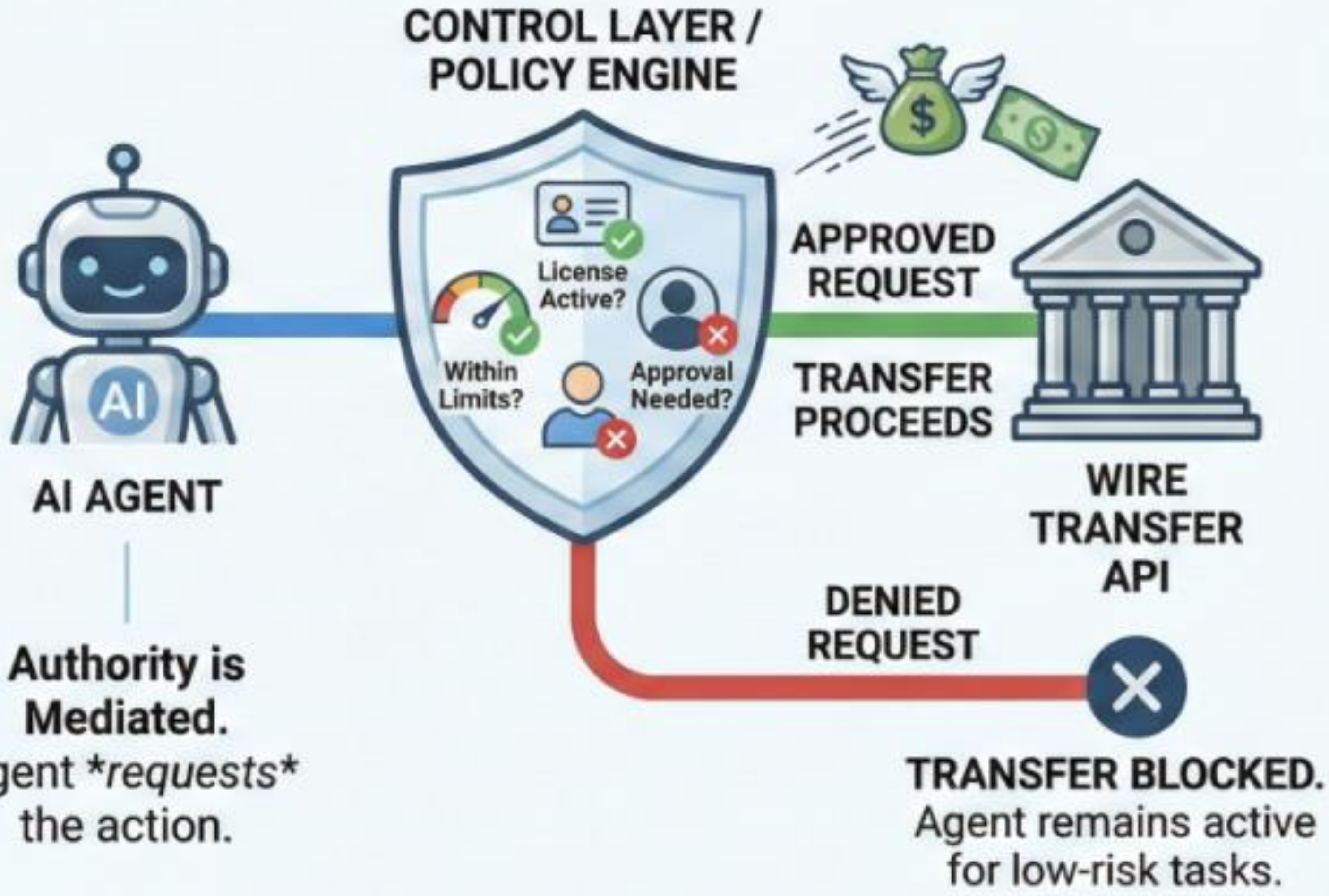
Focus Experts Where It Matters, Automate the Rest.


EXPERTS FOCUS ON UNDERSTANDING, QUANTIFYING, MONITORING, & CONTROLLING RISK
DEEP ANALYSIS & CRITICAL JUDGMENT
UNDERSTAND & QUANTIFY
MONITOR & CONTROL
Verified
ADVANCED VALIDATION
HUMAN-IN-THE-LOOP OVERSIGHT
AUTOMATED RISK TIERING GATE
HIGH-RISK AI MODEL
HIGH RISK LANE (HIGH-TOUCH, EXPERT-LED)
LOWER RISK LANE (LOW-TOUCH, FULLY AUTOMATED)
EFFICIENCY & SPEED AT SCALE
AUTOMATED DOCUMENTATION, TESTING, REGULATORY CHECKS, VALIDATION, & MONITORING
AUTO-DOCUMENTATION
PASS
AUTO-TESTING & VALIDATION
COMPLIANT
REGULATORY & POLICY CHECKS
ACTIVE
CONTINUOUS AUTO-MONITORING
GO
MINIMAL HUMAN EXPERTISE

Key Benefit: Redesigning for risk tiering ensures expert resources are concentrated on high-stakes models, while low-risk AI scales with near-zero manual effort.

EXTERNALIZED AUTHORITY ARCHITECTURE

(Mediated Control)



CONTROL LAYER / POLICY ENGINE

AI AGENT

Authority is Mediated.
Agent *requests* the action.

APPROVED REQUEST

TRANSFER PROCEEDS

WIRE TRANSFER API

DENIED REQUEST

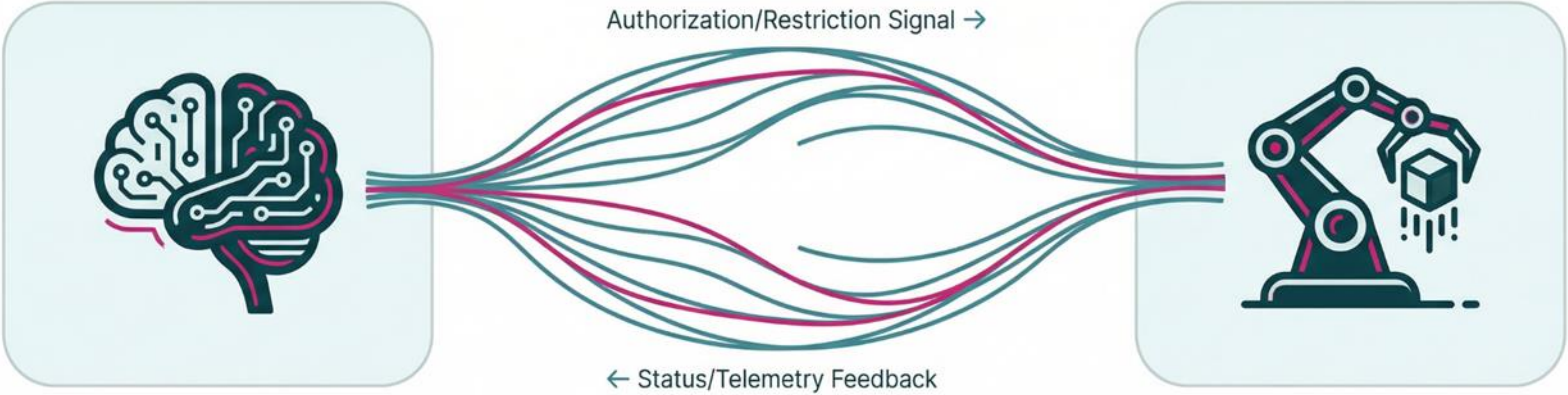
TRANSFER BLOCKED.
Agent remains active for low-risk tasks.

THE NERVOUS SYSTEM OF AI CONTROL: A GOVERNANCE MODEL

THE BRAIN:
AI Governance Hub
(Policy, Risk, Decision)

THE NERVOUS SYSTEM:
Model Context Protocol (MCP)
(Signal, Control, Revocation)

THE MUSCLE:
AI Agent
(Action, Execution, Tool Use)



Authorization/Restriction Signal →

← Status/Telemetry Feedback

Real-time, dynamic control, not just static rules.

THE RED QUEEN

When AI governance becomes selection pressure

Blind Validation: Testing should include probes that models cannot easily distingush distinguish from real user interaction.

Adversarial Co-evolution: Monitoring systems cannot remain static. Safety monitors that learn trrd adapt alongside models are more likely to detect emerging behaviors that fixed rule sets miss.

Honeypots: Governance frameworks should create environments that reward transparency.

VALIDMIND®

Evolution of AI Governance

A tiered framework for AI risk governance

THE KEY CONCEPTUAL SPINE

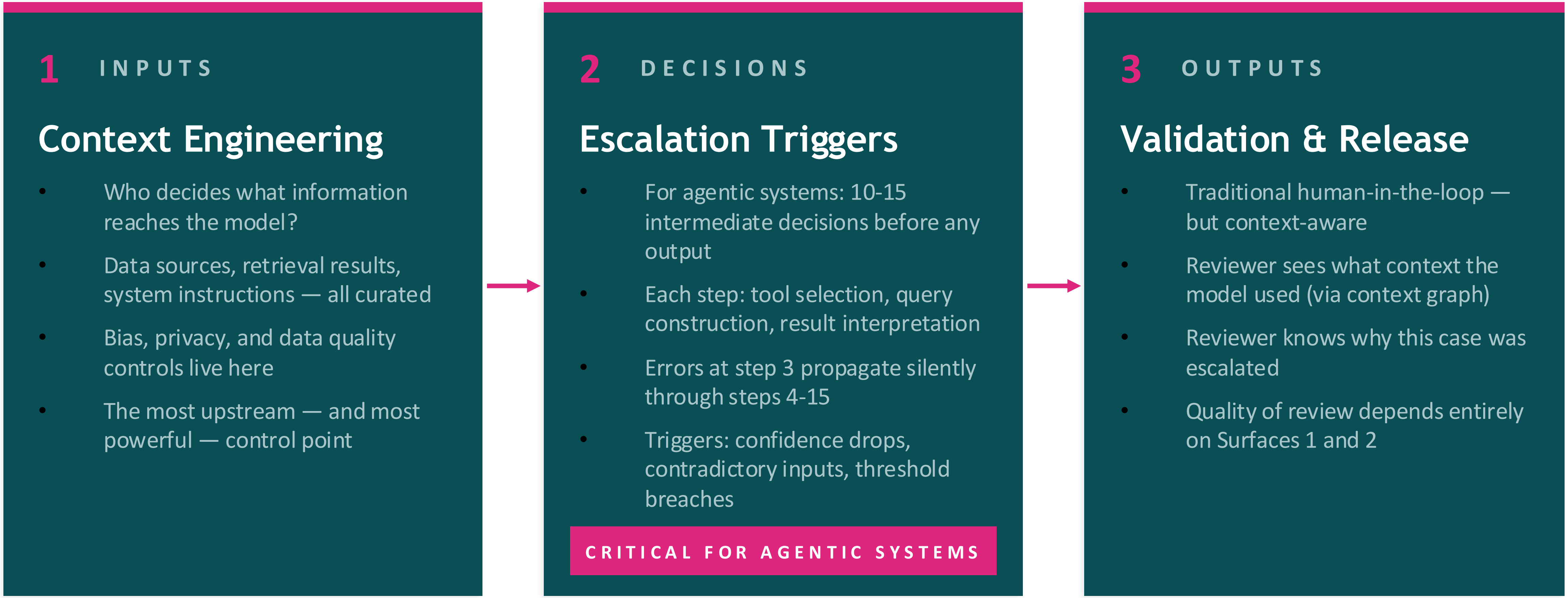
What changes fundamentally at each stage?

Stage	What's Governed	Output Type	Human Role
Traditional ML	A model	Bounded prediction	Validates & monitors
GenAI / LLM	A generation system	Infinite, open-ended	Reviews & corrects
Agentic AI	An autonomous actor	Real-world actions	Intervenes (if aware)

Key insight: Each tier represents a fundamental shift — from governing predictions to governing generation to governing autonomous action.

Human Oversight: Three Governance Surfaces

Context engineering shifts governance from reviewing outputs to designing the information architecture around the model



1 INPUTS

Context Engineering

- Who decides what information reaches the model?
- Data sources, retrieval results, system instructions — all curated
- Bias, privacy, and data quality controls live here
- The most upstream — and most powerful — control point

2 DECISIONS

Escalation Triggers

- For agentic systems: 10-15 intermediate decisions before any output
- Each step: tool selection, query construction, result interpretation
- Errors at step 3 propagate silently through steps 4-15
- Triggers: confidence drops, contradictory inputs, threshold breaches

CRITICAL FOR AGENTIC SYSTEMS

3 OUTPUTS

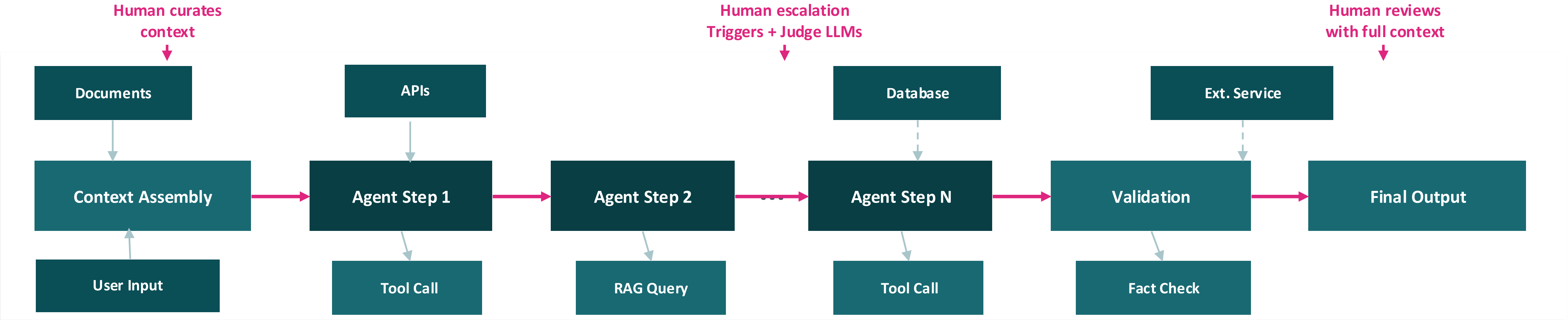
Validation & Release

- Traditional human-in-the-loop — but context-aware
- Reviewer sees what context the model used (via context graph)
- Reviewer knows why this case was escalated
- Quality of review depends entirely on Surfaces 1 and 2

The insight: Most governance frameworks focus on the output. But by the time you see the output, the most consequential decisions have already been made.

The Context Graph as Governance Architecture

The connective tissue across all three governance surfaces — from inputs through decisions to outputs



01

Lineage & Traceability

Every node is auditable. Trace any output back through the chain of context and decisions that produced it. Full provenance from data source to action.

02

Escalation Trigger Points

Nodes where conditions fire human escalation: confidence drops, contradictory inputs, missing context, threshold breaches. Catch problems at step 4, not step 15.

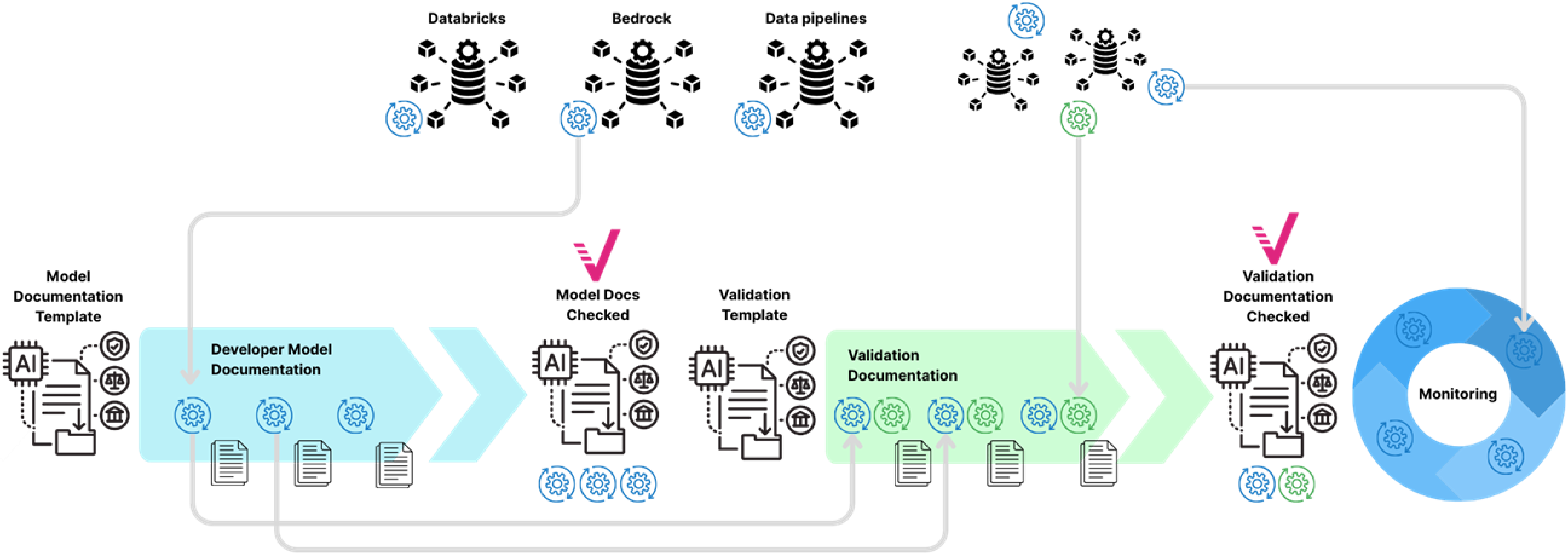
03

Completeness & Quality Gates

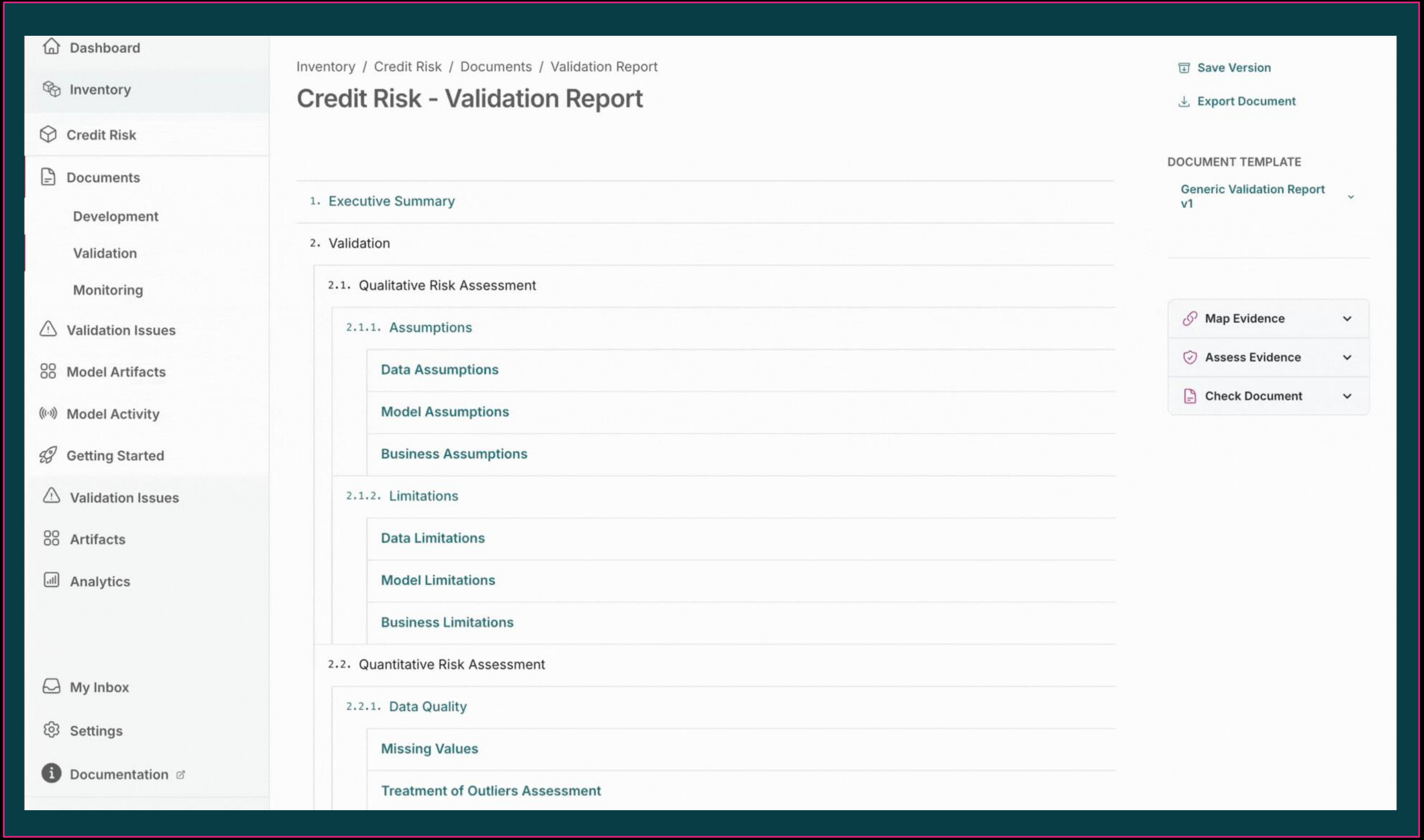
The graph reveals gaps: what context should have been included but wasn't? What sources were consulted vs. available? Governance by design, not by accident.

The bottom line: The context graph isn't just an engineering tool — it's the auditable artifact that makes human oversight meaningful rather than performative.

Automation through Integration



Validation



Inventory / Credit Risk / Documents / Validation Report

Credit Risk - Validation Report

1. Executive Summary

2. Validation

2.1. Qualitative Risk Assessment

2.1.1. Assumptions

Data Assumptions

Model Assumptions

Business Assumptions

2.1.2. Limitations

Data Limitations

Model Limitations

Business Limitations

2.2. Quantitative Risk Assessment

2.2.1. Data Quality

Missing Values

Treatment of Outliers Assessment

Auto-links developer and validator evidence

Assesses the evidence and generate and evaluation

Autogenerates findings based on assessments

Checks the quality of the final validation document against policies

Validator can curate and override every step of the process

VALIDMIND

Model Context Protocol & Governance

How MCP operationalizes the three governance surfaces — and why it changes the oversight model

THE FUNDAMENTAL SHIFT

	Traditional API	MCP
Integration	Developer hardcodes calls	Model discovers & composes
Orchestration	Fixed, deterministic	Dynamic, context-driven
Combining services	Custom glue code per pair	Composed on the fly
Who decides	Developer is architect	Model is architect
Predictability	High — you know what runs	Lower — dynamic decisions

MCP ACROSS THE THREE SURFACES

1 Context (Inputs)

MCP servers expose structured data sources. The model queries "what can you do?" and assembles context from multiple servers dynamically — but every source is declared and traceable.

2 Decisions (Processing)

Every MCP tool call is an auditable intermediate decision. The protocol logs which tool, what parameters, what was returned. These are the nodes in your context graph.

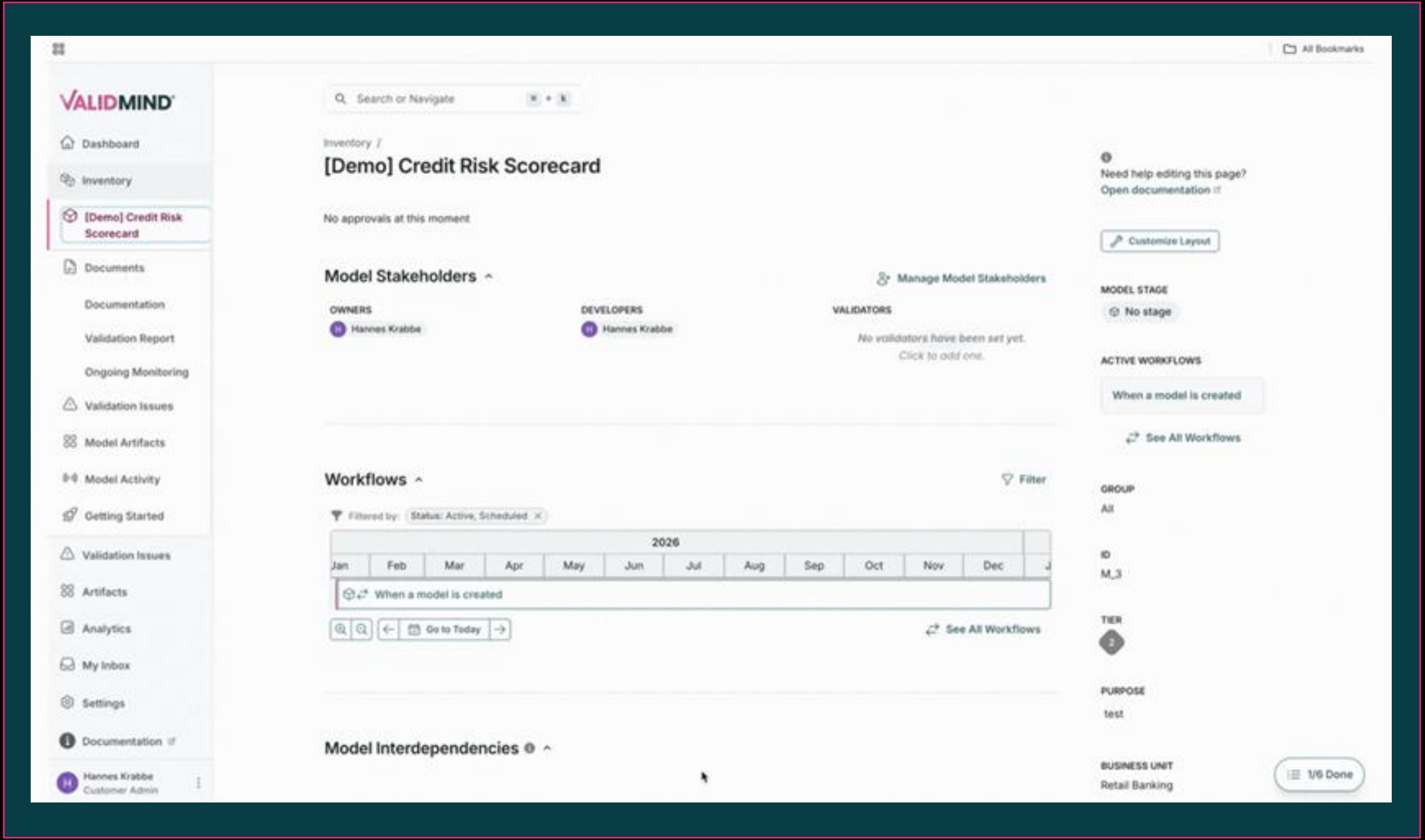
3 Outputs (Actions)

MCP's capability manifest defines what actions are possible. Governance controls can restrict which tools are available, creating hard boundaries on what the model can do.

THE GOVERNANCE TENSION

MCP makes AI systems more capable — **and harder to govern.** With fixed APIs, you know exactly what will be called and when. With MCP, the model makes dynamic decisions about tool use. This is precisely why context graphs, escalation triggers, and structured oversight become non-negotiable. The protocol provides the observability infrastructure — but only if you build governance around it.

Documentation



Inventory /

[Demo] Credit Risk Scorecard

No approvals at this moment

Model Stakeholders

OWNERS: Hannes Krabbe

DEVELOPERS: Hannes Krabbe

VALIDATORS: No validators have been set yet. Click to add one.

Workflows

Filtered by: Status: Active, Scheduled

2026											
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
When a model is created											

Model Interdependencies

Need help editing this page?
Open documentation

MODEL STAGE: No stage

ACTIVE WORKFLOWS: When a model is created

GROUP: All

ID: M_3

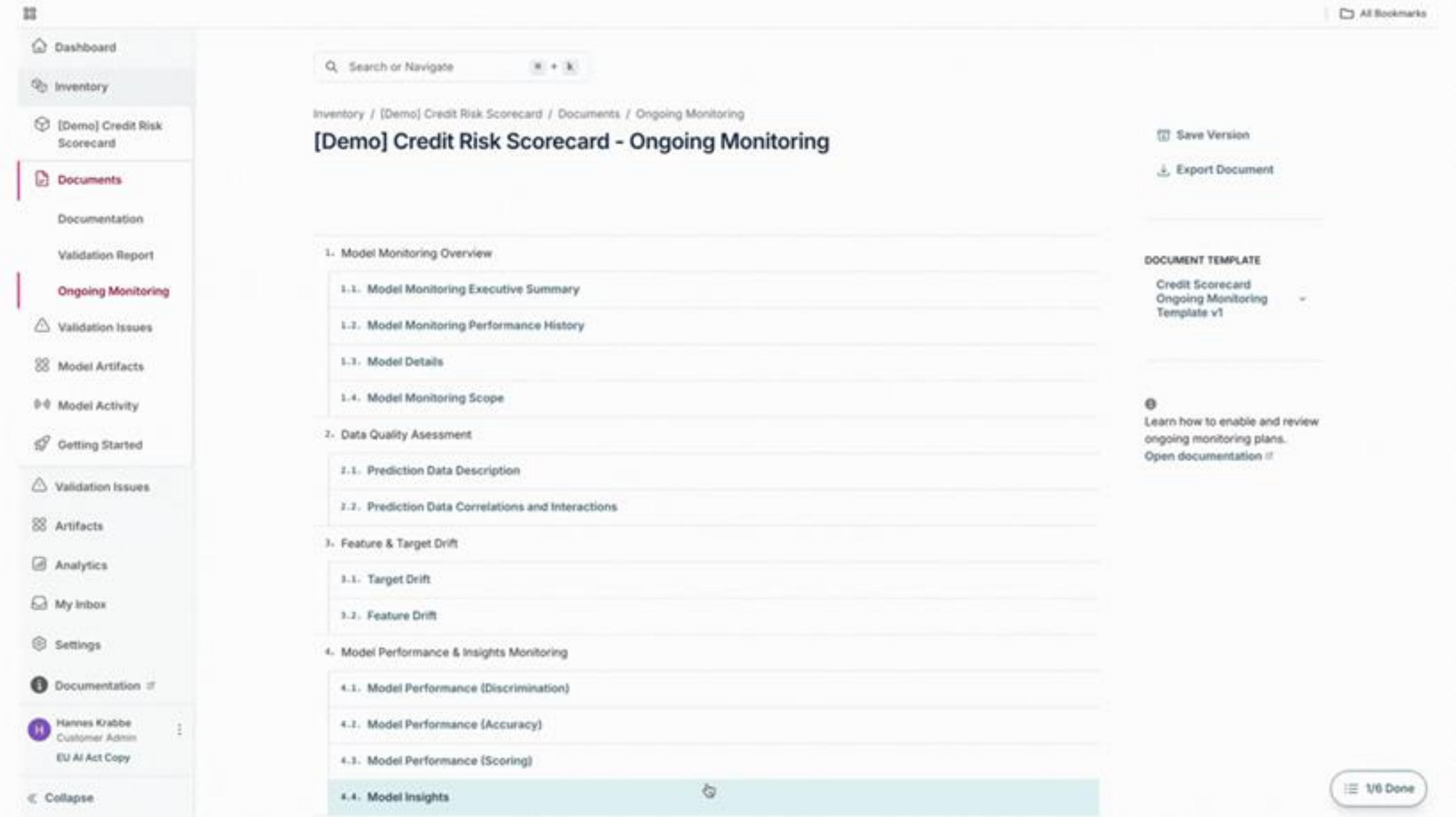
TIER: 2

PURPOSE: test

BUSINESS UNIT: Retail Banking

- Auto-generates required documentation
- Compliance validation checks
- Ensures alignment with legal obligations
- Reduces manual documentation effort
- Improves audit defensibility

Monitoring



Inventory / [Demo] Credit Risk Scorecard / Documents / Ongoing Monitoring

[Demo] Credit Risk Scorecard - Ongoing Monitoring

1. Model Monitoring Overview
   1.1. Model Monitoring Executive Summary
   1.2. Model Monitoring Performance History
   1.3. Model Details
   1.4. Model Monitoring Scope
2. Data Quality Asessment
   2.1. Prediction Data Description
   2.2. Prediction Data Correlations and Interactions
3. Feature & Target Drift
   3.1. Target Drift
   3.2. Feature Drift
4. Model Performance & Insights Monitoring
   4.1. Model Performance (Discrimination)
   4.2. Model Performance (Accuracy)
   4.3. Model Performance (Scoring)
   4.4. Model Insights

DOCUMENT TEMPLATE

Credit Scorecard Ongoing Monitoring Template v1

Learn how to enable and review ongoing monitoring plans.
Open documentation

Continuous performance tracking

Drift detection and alerts

Compliance posture updates

Operational risk signal monitoring

Trending & reporting dashboards

TIER 1

Traditional ML

The baseline of competence — mature frameworks exist (SR 11-7, model inventories, independent validation)

1

Model Performance

Accuracy, drift, overfitting — all variations of the same core concern

2

Data Quality & Bias

Lineage, fairness, representativeness

3

Explainability & Auditability

Can you validate and explain decisions?

This tier is the baseline of competence, not the focus of alarm.

TIER 2

GenAI / LLM

Five genuinely novel risk categories

01 **Output Integrity**

Hallucination — not inaccuracy, but unverifiable open-ended outputs with no ground truth

02 **Prompt Vulnerability**

Injection, adversarial manipulation — no equivalent in traditional ML

03 **IP, Privacy & Compliance**

Copyright, PII leakage, regulatory uncertainty

04 **Human Over-Reliance**

Automation bias at scale — people stop being the check

05 **Vendor Dependency**

Foundation model providers can change models under you silently

Key distinction: The problem isn't "accuracy" in the classical sense — it's unverifiable, open-ended outputs with no ground truth. That's the genuinely new challenge.

TIER 3

Agentic AI

Five categorical leaps — each a step-change from the prior tier

01 Irreversibility & Real-World Consequence

Actions, not outputs — decisions that cannot be undone

02 Cascade & Compounding Errors

Multi-step failure propagation through complex chains

03 Accountability & Auditability Gaps

Who is responsible when an agent chain fails?

04 Reduced Human Oversight by Design

The whole point of agents is to remove humans from the loop

05 Emergent Multi-Agent Behavior

Agent-to-agent interactions nobody designed or anticipated

The paradigm shift: We move from governing what a model says to governing what an autonomous system does — actions with real-world, irreversible consequences.