SPOTLIGHT ON AI MODEL VALIDATION

AI Model Validation: Risk to Reliability through Quality Assurance Standards

A Spotlight on Shifting Risk to Reliability through Quality Assurance Standards, Risk Management, and Compliance in AI Model Validation

StratisAdvisory

RISK | STRATEGY | CYBER | COMPLIANCE MANAGEMENT

Organizational Value of Model Validation

Model validation is essential to ensure accuracy of monitoring and detection tools that aid in ensuring regulatory or contractual compliance. This should consider data quality, governance, performance testing, and overall oversight to reduce false positives and negatives. The introduction of AI-driven models enhance efficiency in processing large, complex datasets and aid in error reduction, utilizing preset logic and rules against risk scenarios tailored to the organization's business profile. Potential gaps to AI-driven models should be considered, such as how alerts are generated, data bias that can distort outcomes, and necessary recalibration with new inputs.

Validation Methodology

Data Validation → Processing → Reporting

Core Elements

Evaluation of Conceptual Soundness
Review of documentation and empirical evidence that supports the methods used and the variables selected for the model.

Outcomes Analysis
Compares the model outputs to observed actual outcomes. Defines expected outcome ranges relative to objectives.

Ongoing Monitoring
Review changes in products, exposures, activities, clients, or market conditions for potential scope expansion.

AI Risk Management Framework: From Risk to Reliability

AI Model Validation is the independent assessment of models used to ensure they are accurate, reliable, and aligned with industry best practice. The National Institute of Standards and Technology (NIST) introduced the AI Risk Management Framework (AI RMF), outlining core functions that guide institutions with oversight, identify risks, thresholds setting, and continuously monitoring. Alongside these functions are the characteristics of trustworthy AI which serve as parameters against which models must be validated.

Govern

Ensures accountability and oversight for both parameters and risk management.

Manage

Focuses on mitigation, monitoring, and incident response when models fall outside parameters.

Map

Identifies risks tied to those parameters, such as bias, drift, misuse, or systemic vulnerabilities.

Measure

Defines how organizations establish metrics and thresholds for accuracy and bias.

- **Valid and Reliable:** Consistent performance as intended across contexts and time.
- **Explainable and Interpretable:** Decisions can be understood and justified by regulators, auditors, users.
- **Accountable and Transparent:** Clear responsibility structures exist for oversight, validation, and decisions.
- **Fair with Harmful Bias:** Models deliver equitable outcomes across demographics.
- **Privacy Enhanced:** Sensitive data is protected through privacy-by-design principles.

Human-in-the-loop vs. Pure AI Logic

Quality in AI model validation is anchored on how outcomes are generated, whether through logic or supported by human oversight. One theme stands out: quality in AI model validation cannot rely on automated logic alone. Human-in-the-loop is vital to ensure accountability, reliability, and compliance. This view further shifts quality assurance to a more comprehensive compliance strategy shifting risks into reliability.

Primary Themes in Quality

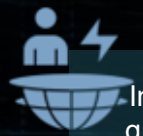
Quality anchored in measurable parameters, measurable thresholds, bias audits, structured testing, and continuous monitoring.

AI systems should have defined tasks, scope, and implementation method, while documenting its limits and oversight guidance.

Governance frameworks ensuring structured oversight with executive management or board-level accountability and supervisory review.

Responsibility remains with organizations and individuals, not delegated solely to automated systems. Ensure quality of AI outcomes are remain socially responsible.

Human-in-the-loop anchors accountability, reinforcing transparency and risk management embedded in quality assurance, ensuring processes are documented and auditable.

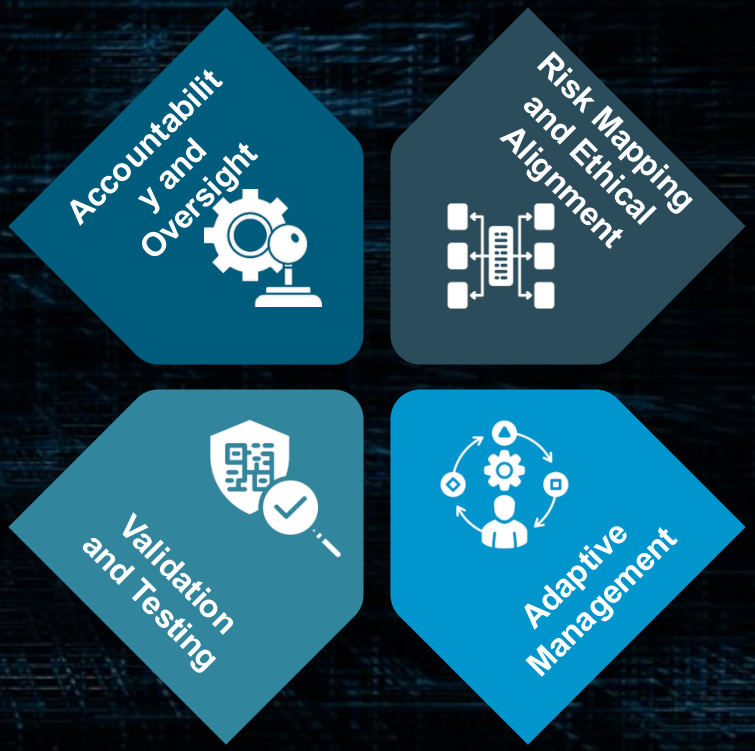


Independent validation within governance strengthens trust, aligning with the organization's risk appetite and meet examiner expectations, if applied in a regulated field.

Path Forward: Building an AI Model Validation Framework

Building a framework through embedded parameters from NIST create a unified charter where quality is not compromised by speed and oversight is not sacrificed for automation. An ideal framework transforms blockers into solutions and converge to deliver statutory-ready assurance.

- Drawn from NIST's Govern function, this focuses on clear accountability at board level, separation of duties, and independent validation teams.
- Human oversight is built into the loop to ensure models are not left to operate unchecked.
- Accountability layer closes governance gaps and prevents analyst bottlenecks from undermining trust.



- Derived from NIST's Map function, this layer ensures risks are identified before deployment. It requires cataloging models, mapping potential harms, and embedding fairness principles into design.
- By aligning models with ethical and supervisory expectations, this layer combats bias, opacity, and fairness gaps, ensuring institutions build trust into their systems from the

- This component integrates NIST's Measure function, transparency in ISO standards, and ECB validation rules. It emphasizes bias audits, robustness checks, performance benchmarking, and replicable documentation.
- Supervisors must be able to reproduce model outcomes, making documentation a core requirement and undergo quarterly audits and performance checks.

- This is anchored in NIST's Manage function which emphasizes resilience and adaptability. It focuses on monitoring drift, automating workflows, and integrating analyst feedback into continuous learning loops.
- Agentic AI approaches can be applied to automate regulator-ready narratives and scale monitoring while keeping humans in the loop

Discover what Stratis Advisory can do for you

AI Program Development

As Agentic AI capabilities—such as autonomous reasoning, planning, and self-refining—become embedded in operations, an AI program ensures alignment with regulatory expectations and mitigates emerging risks. At full spectrum, from design, deployment, up to governance and assurance, Stratis can help you develop a scale-appropriate AI program tailored to your operating model, risk profile, and jurisdictional exposure ensuring your Agentic AI systems remain compliant and future-proof.

AI Deployment and Assurance

The AI agent assurance process is crucial to build confidence in the operational trustworthiness, safety and responsibility of an AI agent throughout its lifecycle, from pre- to- post deployment. Stratis can help perform a structured approach in your agent assurance, from the definition of objectives and AI agent performance metrics, testing and evaluation, and continuous monitoring and calibration, to documentation and reporting and independent audits. This will ensure your AI agent's reliability and compliance to maintain its performance and manage risks.

AI Governance

AI governance sets the parameters for accountability, systems, policies and controls, to maintain a safe, ethical and compliant AI agent model. Stratis can assist you in developing and executing your AI governance, lifecycle controls and regulatory readiness, which will ensure that your AI agents remain under human and systemic oversight.

LAUNCH | SCALE | OPTIMIZE

Stratis Advisory | 1.415.352.1060 | www.stratisadvisory.com | info@stratisadvisory.com