FROM STATISTICAL MODELS TO GENERATIVE SYSTEMS: CLOSING THE GOVERNANCE GAP

AI Model Risk Management & MLOps Governance

Executive briefing — Chief Risk Officer & Model Risk Leadership | ~20 min

Why This Matters Now

- Generative AI and LLM-based systems are moving into production faster than governance frameworks can adapt
- Existing model risk programs were built for statistical, credit, and pricing models with stable, well-defined behavior
- Business units are adopting AI copilots, chatbots, and decision-support tools outside formal model inventories
- Regulators and internal audit are beginning to ask questions existing frameworks were not designed to answer
- This briefing outlines the gaps and a practical path to close them

Where Traditional Model Risk Frameworks Fall Short

- Traditional frameworks (e.g., SR 11-7-style guidance) assume a defined input-output function with measurable, stable error rates
- Generative models produce open-ended, non-deterministic outputs that resist conventional back-testing
- Validation methods built for regression and classification models do not test for hallucination, tone, or reasoning failures
- Third-party foundation models are often black boxes: no access to training data, weights, or full documentation
- Model boundaries blur when prompts, retrieval systems, and fine-tuning layers change behavior post-deployment

Expanded Risk Taxonomy for AI Systems

- Hallucination: confident, fabricated, or unsupported outputs presented as fact
- Prompt injection: malicious or unintended instructions embedded in inputs, documents, or retrieved content
- Data leakage: sensitive or proprietary information exposed through outputs, logs, or third-party model providers
- Model and data drift: performance degradation as usage patterns, prompts, or underlying model versions change
- Third-party and vendor risk: dependency on external model providers with limited transparency or control

Model Inventory and Registration

- Every AI system in use — including embedded vendor features and internal copilots — must be captured in a central inventory
- Registration should capture: purpose, owner, data sources, model provider/version, and business criticality tier
- Shadow AI usage (unsanctioned tools adopted by business units) is a primary near-term inventory gap
- Inventory should distinguish traditional models, generative models, and hybrid/agentic systems, since risk profiles differ
- Criticality tiering determines the depth of validation, monitoring, and approval required before deployment

Validation and Pre-Deployment Testing

- Extend validation beyond accuracy metrics to include robustness, adversarial testing, and prompt injection resistance
- Red-teaming and structured adversarial testing should be standard practice for customer-facing generative systems
- Fairness and bias testing must account for language, tone, and demographic proxies, not only structured features
- Human-in-the-loop review is required for high-criticality use cases prior to go-live
- Sign-off should be tied to intended use case, not the underlying model in isolation — the same model can carry different risk in different applications

Ongoing Monitoring Post-Deployment

- Performance monitoring must run continuously, not only at periodic revalidation checkpoints, given how quickly usage and prompts evolve
- Drift monitoring should track both data drift and behavioral drift from underlying vendor model updates outside the organization's control
- Fairness and output-quality monitoring should sample production interactions, not rely solely on pre-deployment test sets
- Automated guardrails (content filters, output validators, escalation triggers) should be paired with periodic manual review
- Monitoring ownership should be explicit and tied to the three-lines-of-defense structure, not left to the deploying business unit alone

Three Lines of Defense Applied to AI

- First line: business and technology teams building and operating AI systems, accountable for day-to-day controls
- Second line: model risk and compliance functions setting standards, performing independent validation, and monitoring adherence
- Third line: internal audit providing independent assurance over the end-to-end governance program
- AI systems require closer coordination between lines given the speed of vendor model updates and shorter development cycles
- Second-line capacity and AI-specific expertise are typically the binding constraint in scaling this model

Vendor and Third-Party Model Risk Assessment

- Third-party model risk should be assessed with the same rigor as internally developed models, adjusted for reduced transparency
- Key diligence areas: training data provenance (to the extent disclosed), update cadence, data handling, and sub-processor use
- Contractual protections should address model change notification, data usage restrictions, and audit or attestation rights
- Concentration risk matters: heavy reliance on a single foundation model provider is itself a risk to be tracked and reported
- Exit and fallback plans should be defined before deployment, not after an incident

Illustrative Governance Rollout Example

- Illustrative scenario, not a verified case study — presented to show a plausible sequencing, not an audited outcome
- Phase 1 (0-3 months): inventory sweep, criticality tiering, and interim policy for new AI use cases
- Phase 2 (3-6 months): validation standards and monitoring tooling stood up for highest-tier use cases first
- Phase 3 (6-12 months): three-lines-of-defense roles formalized, vendor assessment process embedded in procurement
- Phase 4 (12+ months): steady-state monitoring, periodic revalidation cadence, and audit readiness

Regulatory Context

- Existing model risk guidance such as SR 11-7 remains the foundational reference point and generally extends to AI systems in principle
- Supervisory expectations increasingly emphasize explainability, monitoring, and accountability for automated decision-making generally
- AI-specific regulation is still developing across jurisdictions, with more prescriptive requirements expected over time
- Approach: treat current guidance as a floor, and build governance capacity ahead of specific new AI regulatory requirements
- Legal and compliance should track jurisdiction-specific developments relevant to the firm's footprint on an ongoing basis

Incident Response for Model Failures

- Define what constitutes an AI incident: harmful output, data exposure, biased decision, or system misuse
- Establish clear escalation paths from first-line teams to model risk, compliance, and legal within defined time windows
- Maintain a kill-switch or rollback capability for production AI systems, tested prior to go-live
- Post-incident review should feed back into validation standards and monitoring thresholds, not remain a one-time fix
- Track incident trends over time as a leading indicator of program maturity, not just individual event severity

Documentation and Audit Trail Requirements

- Every registered model requires a documented purpose, validation record, approval history, and monitoring log
- Prompt and output logging should be retained for high-criticality use cases to support review and investigation
- Version control is required for prompts, fine-tuning data, and configuration changes, not only for the base model
- Documentation should be structured for both internal audit and external examiner review without ad hoc reconstruction
- Retention and access policies must balance audit readiness with data minimization and privacy requirements

Common Pitfalls to Avoid

- Treating AI governance as a one-time compliance exercise rather than a continuously monitored program
- Allowing business units to adopt AI tools outside the inventory process, creating shadow AI exposure
- Applying pre-generative-AI validation checklists without adapting them to open-ended model behavior
- Underestimating dependency risk on third-party model providers with limited visibility into changes
- Delaying second-line and audit involvement until after significant deployment, rather than building it in from the start

Next Steps and the Ask

- Approve a 90-day inventory sweep to establish a complete, tiered view of AI use across the organization
- Commit second-line resourcing to build AI-specific validation and monitoring standards for highest-tier use cases
- Mandate that new AI use cases route through registration and approval before production deployment, effective immediately
- Direct procurement and legal to embed AI vendor risk assessment into the existing third-party risk process
- Establish a quarterly reporting cadence to this group on inventory growth, validation backlog, and incident trends