Security And Governance Playbook

Suganya Purushothaman

AI Product Manager | Governance, Security & Responsible AI

AI/ML Security & Governance — 10 Practical Steps

1. Data Minimization Collect and store only the minimum data necessary for the model. Remove or mask personally identifiable information (PII) before ingestion. This reduces privacy risks and compliance overhead.

2. Access Control & Audit Logs Restrict access to sensitive datasets and model endpoints. Maintain immutable audit logs to track who accessed what and when, supporting accountability and investigations.

3. PII Filtering & Anonymization Before sending data to external APIs or cloud services, apply PII filters and anonymization. Replace sensitive tokens with pseudonyms to safeguard user identities.

4. Model Explainability Document why a model was chosen, its inputs, and its decision-making process. Provide simple explainability notes for business stakeholders to build trust.

5. Secure Inference Ensure inference requests are encrypted in transit (HTTPS/TLS). Use private endpoints, authentication, and rate-limiting to prevent misuse of your models.

6. Bias & Fairness Checks Continuously test datasets and model outputs for bias. Track fairness metrics and publish mitigation plans as part of release notes to ensure accountability.

7. Governance Gate for Releases Introduce a formal release checklist. Each model version must pass governance checks: data quality, bias, privacy, reproducibility, and cost thresholds.

8. Human-in-the-loop & Escalation For high-risk use cases (e.g., healthcare, finance), involve human reviewers before final decisions. Define clear escalation workflows for sensitive outputs.

9. Monitoring & Drift Detection Deploy monitoring for input distribution shifts and performance decay. Trigger retraining, alerts, or rollbacks when drift or anomalies are detected.

10. Contracts & Third-party Assessment When integrating external APIs or third-party models, vet them for compliance and add clear SLA/security clauses. Ensure vendors meet your governance standards.