GLOBAL CYBER RISK SERVICES

Beyond the Algorithm

Building Trust, Security, and Accountability in AI

ALVAREZ & MARSAL

LEADERSHIP. ACTION. RESULTS.℠

Executive Summary

AI has moved beyond experimentation and efficiency to the core of competitive strategy. While India Inc. is investing aggressively in intelligent systems, few are matching those investments with security and governance. As enterprises race to deploy generative AI, predictive analytics, and autonomous systems, a critical gap is emerging: **AI infrastructure is outpacing AI security and governance.**

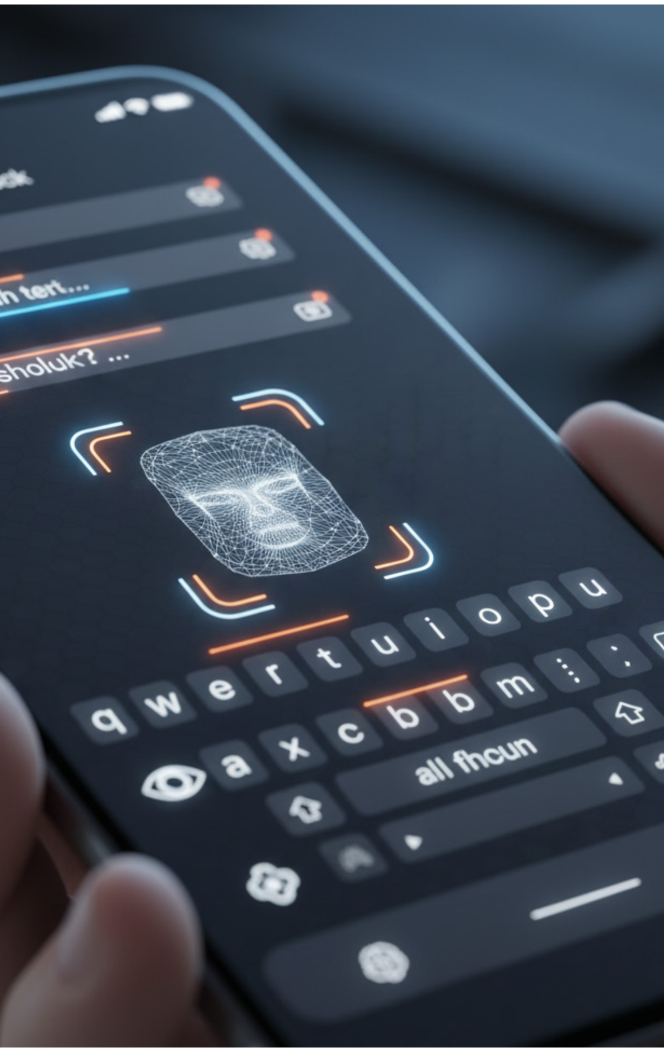
Models trained on sensitive data, APIs exposed to adversarial attacks, and algorithmic decisions that lack accountability are creating cascading risks—from data breaches and model poisoning to regulatory penalties and reputational collapse.

This report is based on a month-long survey of senior executives to assess the maturity of AI security and governance practices in India. It explores the evolution of AI adoption with a focus on the critical pillars of governance, responsible AI, data and model development security, model deployment and operationalization, monitoring, incident reporting and compliance.

The survey targeted India's senior technology, security, and risk leaders, including Chief Information Security Officers (CISOs), Chief Information Officers (CIOs), Chief Technology Officers (CTOs), and Chief Risk Officers (CROs) from enterprises of varying sizes across finance, technology, healthcare, manufacturing, retail, and other key industries.

The survey examined organizations' progress in AI adoption, governance maturity, responsible AI initiatives, and end-to-end security implementation. Executives provided valuable insights into their deployment strategies, risk management approaches, and preparedness for navigating the evolving regulatory landscape.

The findings reflect a cross-sectional view of AI readiness and challenges in the Indian enterprise landscape, providing a data-driven foundation for strategic decision-making and policy development.



A directional summary is provided below.

1. **AI Adoption: Momentum in Leading Sectors, Experimentation in Others**
   Large enterprises within the Banking, Financial Services, and Insurance (BFSI) and Technology, Media, and Telecommunications (TMT) sectors are scaling AI adoption, while others remain in early stages, relying on isolated use cases and hybrid implementations to balance scalability, cost, and control.

2. **Governance Crisis: Adoption is Outpacing Oversight**
   Rapid AI deployment significantly exceeds governance maturity—only a minority have structured frameworks, comprehensive risk assessments, real-time monitoring, or AI-specific incident response capabilities.

3. **Responsible AI: Acknowledged but Under-Delivered**
   Despite widespread recognition, responsible AI shows significant implementation gaps in explainability, bias detection, data privacy, and model integrity, creating ethical and regulatory challenges.

4. **Lifecycle Security: Foundational Controls in Place, Advanced Capabilities Still Nascent**
   While foundational security controls exist—such as access controls and secure environments—more advanced defenses like red teaming, prompt-injection protection, hallucination detection, along with post-deployment oversight mechanisms such as continuous monitoring, AI incident response, and formal audits, remain significantly underdeveloped.

Key Insights from the Survey

1 AI Adoption is Accelerating but Still Fragmented

- Only 15% of organizations have deployed AI extensively across business units and customer touchpoints.
- 48% use hybrid implementation models, combining software-as-a-service (SaaS), original equipment manufacturer (OEM), and custom-built solutions—indicating a phase of experimentation and optimization.

2 Governance Maturity Remains Low

- While 60% have basic AI governance policies, only 19% have conducted structured risk assessments covering legal, ethical, and societal dimensions.
- 81% reported partial or no control over AI monitoring and governance metrics.
- Just 22% conduct AI-specific due diligence for third-party vendors, exposing them to unmanaged external risks.

3 Responsible AI is Recognized but Rarely Operationalized

- Fewer than 20% of organizations have implemented explainability, bias detection, or fairness mechanisms in their AI systems.
- 60% lack formal model integrity assurance processes.
- 51% have no controls to detect or mitigate AI hallucinations, and only 15% have AI-specific incident response plans.
- 66% conduct no formal audits or rely on ad hoc checks, undermining regulatory readiness.

4 Security Practices are Foundational but Not Advanced

- 52% have secured development environments with access controls, but fewer than 30% conduct penetration testing or red teaming.
- Only 19% have mechanisms to detect and mitigate data poisoning during model training.
- 56% mandate security reviews before deployment, yet only 26% have controls for prompt-injection attacks, and 30% rely on basic validation.

These insights underscore a critical gap between AI ambition and operational readiness. While foundational steps have been taken—such as securing development environments and implementing access controls—other foundational elements such as governance frameworks, risk assessment, red-teaming, real-time monitoring, and ethical oversight remain significantly underdeveloped.

To succeed in the AI-driven future, organizations must move beyond compliance checklists and embrace a proactive, end-to-end approach to AI governance and security. Embedding transparency, accountability, and ethical stewardship across the AI lifecycle is not only a regulatory requirement—it is a strategic imperative for building trust, resilience, and long-term value.

The Rise of AI Adoption and the Imperative for Responsible Governance

AI adoption across organizations has surged, driven by its potential to enhance efficiency, improve decision-making, and accelerate innovation. From chatbots in banking to predictive maintenance in manufacturing and supply chain optimization in retail, AI is transforming operations across industries. However, this rapid integration also exposes organizations to evolving cybersecurity and ethical risks. Therefore, AI initiatives must strike a balance between innovation and robust security—ensuring systems are secure, trustworthy, and resilient from the outset. AI adoption is primarily led by large enterprises, with 74% of survey respondents in India representing firms with over 1,000 employees. The financial and technology sectors dominate this group, accounting for 66.6% of respondents. Known for their regulatory rigor and advanced digital capabilities, these industries are well-positioned to leverage AI for automation, enhanced risk controls, and innovation—making AI a natural fit for their operational needs.

Modern enterprises deploy AI across diverse functions, each presenting unique security considerations:

Customer-Facing Applications

- Chatbots and Virtual Assistants: Risk of prompt-injection attacks, data leakage, and inappropriate responses
- Recommendation Engines: Vulnerability to manipulation affecting business decisions
- Personalization Systems: Privacy concerns and potential for discriminatory outcomes
- Content Generation: Risk of generating harmful, biased, or copyrighted content

Internal Operations

- Document Processing and Analysis: Risk of sensitive data exposure through training datasets or model outputs
- Code Generation and Review: Potential security vulnerabilities in generated code and intellectual property (IP) leakage
- Automated Decision Systems: Bias in hiring, lending, or resource allocation
- Fraud Detection: Adversarial attacks used to evade detection systems

Mission-Critical Systems

- Autonomous Vehicles and Robotics: Safety-critical failure modes
- Medical Diagnosis and Treatment: Patient safety and regulatory compliance
- Financial Trading and Risk Assessment: Market manipulation and systemic risk
- Infrastructure and Security Operations: SIEM, threat detection, and automated response

Despite this momentum, only 15% of organizations surveyed in India report extensive AI deployment across business units and customer touchpoints. While adoption is increasing, most organizations remain in the early stages, focusing on isolated use cases rather than enterprise-wide integration. This underscores the untapped potential of AI to transform operations and customer experiences at scale.

Implementation strategies vary widely. Some organizations rely on SaaS platforms for rapid deployment, others embed AI through OEM integrations; while a few invest in bespoke models tailored to their unique needs. The most common approach is hybrid—combining SaaS, OEM, and custom-built solutions.

Our survey confirms this: 48% use hybrid implementations, while only 18.5% build bespoke solutions. This diversity reflects an experimental phase as organizations seek the right balance of capability, cost, and strategic fit.

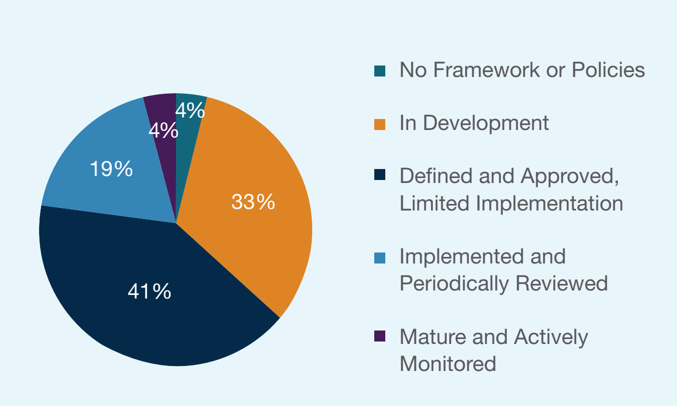
AI Governance: A Strategic Imperative

As AI becomes deeply embedded in business operations, products, and decision-making, establishing a strong governance framework is no longer optional—it is critical. AI governance encompasses the policies, roles, processes, and controls that guide responsible development, deployment, and oversight. It is not merely about compliance; it is a strategic necessity to ensure AI is ethical, secure, and aligned with organizational values and societal expectations.

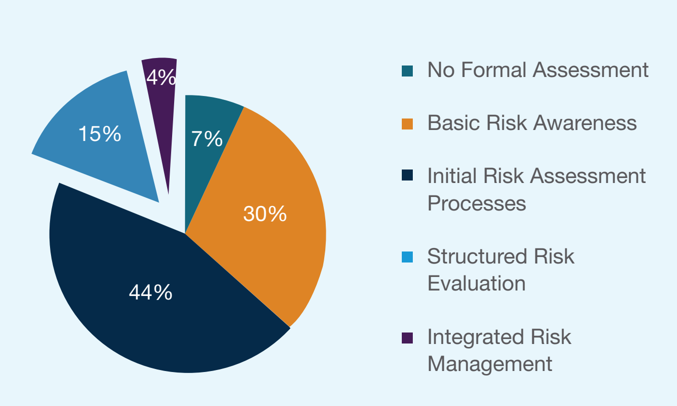
AI introduces distinct risks compared to traditional software due to its probabilistic decision-making, data dependency, and often opaque logic. These characteristics create vulnerabilities such as algorithmic bias, limited explainability, privacy breaches, and unintended systemic consequences. The risks span multiple dimensions—technical failures producing biased outcomes, legal exposure from regulatory violations, and reputational damage from stakeholder backlash. A single poorly governed model can perpetuate discrimination, attract regulatory sanctions, or trigger public controversy. Effective governance frameworks provide a foundation for building and maintaining trust among internal teams and external stakeholders.

Regulatory pressure is intensifying. Frameworks such as the European Union Artificial Intelligence Act (EU AI Act), the Organisation for Economic Co-operation and Development (OECD) principles, and the National Institute of Standards and Technology (NIST- Risk Management Standard 2023) standards demand transparency, accountability, and fairness. Organizations that fail to embed governance, risk delays and penalties, while those that invest early gain a competitive edge.

AI Governance Implementation Status

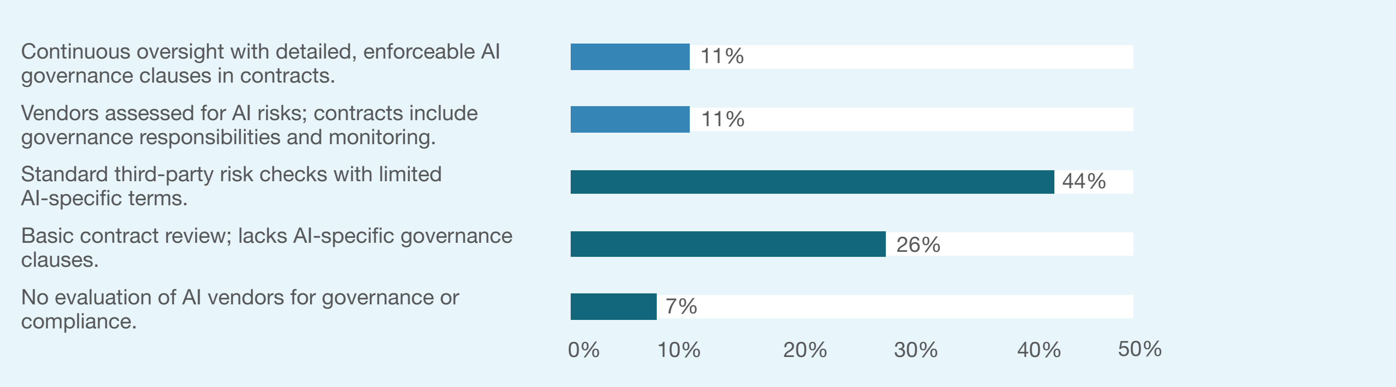


Maturity of AI Risk Management Processes



Despite the urgent need, AI governance maturity remains critically low. While 60% of organizations surveyed in India report having basic governance and security policies, most lack structured implementation and clearly defined roles. AI initiatives typically operate in silos, leading to inconsistent risk assessments and fragmented ethical reviews. Only 19% have conducted comprehensive risk assessments covering legal, ethical, and societal dimensions or integrated AI risks into enterprise governance frameworks. The visibility gap is particularly concerning: 81% of organizations have only partial or no control over AI monitoring and governance metrics, lacking the real-time dashboards and automated alerts that characterize mature governance programs.

Governance Maturity in Managing Third-Party AI Vendors

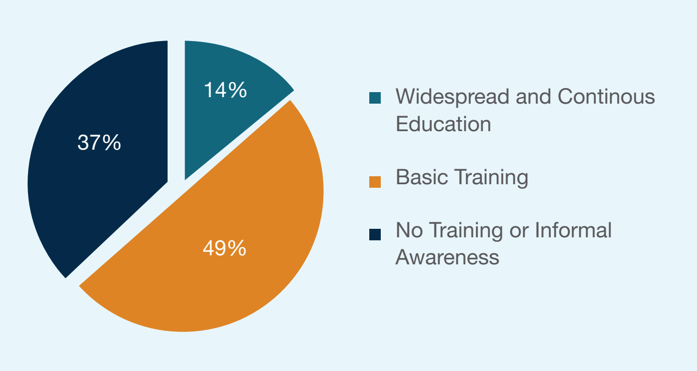


Third-party oversight is also weak. Only 22% conduct AI-specific due diligence for vendors, while 45% rely on traditional processes that fail to address AI-specific risks. This gap is especially critical in hybrid environments where in-house and external AI tools coexist.

Training represents another critical weakness: only 14% of organizations surveyed in India have implemented comprehensive, structured programs featuring role-specific guidance and continuous education embedded in learning frameworks. In stark contrast, 37% provide no AI training at all, while the remaining 49% rely on informal awareness sessions that lack systematic structure, consistent delivery, or ongoing reinforcement mechanisms.

Ultimately, AI governance must evolve in tandem with organizational maturity and regulatory landscapes. While there is no universal model, the principles of transparency, accountability, and trust are foundational. Organizations that embed these principles into scalable frameworks will be better equipped to manage risks and unlock AI's full potential.

Level of training and awareness around AI



Best Practices for AI Governance

Area	Specific Actionable Steps
Governance Policy and Roles	• **Formalize AI governance policy:** Move beyond ad hoc policies; define clear principles aligned with OECD principles, EU AI Act, and NIST standards among others. • **Establish an AI acceptable-use policy:** Define approved tools, prohibited use cases, and data-handling standards; include training requirements, violation reporting, third-party approvals, and boundaries between personal and professional use. • **Integrate AI domains into enterprise policy:** Formally embed AI risk management, model development and deployment, incident response, third-party risk, data governance, and monitoring within the overarching IT policy framework. • **Assign dedicated roles:** Appoint a Responsible AI Officer and/or an AI Product Owner to ensure accountability. • **Create a cross-functional Governance Committee:** Include legal, compliance, risk, cybersecurity, and business leaders for holistic oversight. Survey Insight: 60% have some policy/roles, but structured implementation is limited.
Risk Assessment	• **Conduct structured AI risk assessments:** Cover legal, ethical, societal, and business risks across the lifecycle. • **Integrate AI risks into the enterprise risk framework:** Ensure AI risks are part of overall risk management. • **Use scenario-based testing:** Simulate bias, privacy breaches, and operational failures to validate controls. Survey Insight: Only 19% have completed structured AI risk assessments.

Third-Party Oversight

- **Strengthen vendor due diligence:** Include contractual clauses for bias testing, explainability, and regulatory compliance.
- **Mandate cybersecurity and data controls:** Require robust logging, access controls, and compliance with data-localization obligations where applicable.
- **Implement model-level oversight:** Periodically review vendor models for bias mitigation, explainability, and transparency.
- **Conduct external AI tool assessments:** Independently validate vendor models for fairness, security, and ethical alignment.
- **Conduct continuous monitoring:** Require periodic compliance reports and performance reviews.

Survey Insight: Only 22% conduct AI-specific due diligence; 45% rely on generic processes.

Monitoring and Visibility

- **Deploy real-time governance dashboards:** Track security compliance, accuracy, and compliance metrics.
- **Configure automated alerts:** Configure alerts for anomalies, security, or ethical breaches.
- **Periodic governance reviews:** Validate adherence to policies and update controls.

Survey Insight: 81% have partial or no visibility into AI operations.

Accountability and Transparency

- **Maintain audit trails:** Log all model decisions and governance actions.
- **Enable explainability for stakeholders:** Provide clear, interpretable outputs.
- **Disclose AI use appropriately:** Share high-level governance practices to build trust.
- **Create ethical review boards:** Require approvals for high-risk AI deployments before launch.

Responsible AI: From Ethical Ideal to Business Imperative

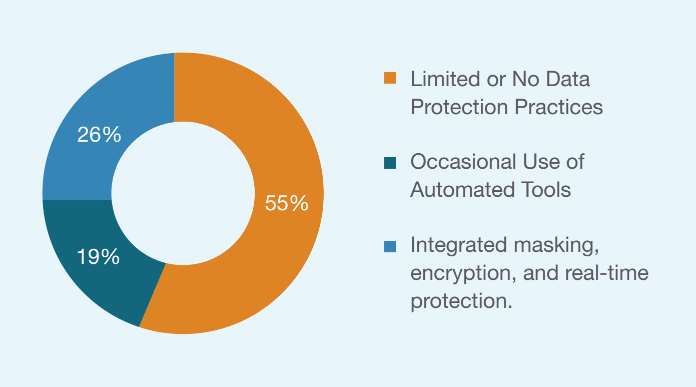
In recent years, responsible AI has made a decisive leap—from ethical ideal to business mandate and regulatory requirement. AI's deep integration across healthcare, finance, government, and consumer services has fundamentally shifted expectations: organizations must now deliver systems that are not just powerful and scalable, but provably fair, accountable, and secure. The 2025 reality is clear: responsible AI has transitioned from nice-to-have to must-have, forming the essential foundation for sustainable innovation and stakeholder trust.

At its core, responsible AI enables organizations to align technological advancements with legal, societal, and stakeholder expectations. These expectations may be codified in law or shaped by evolving cultural norms and values. The growing influence of global frameworks—such as the OECD AI Principles, the EU AI Act, and national standards from bodies like NIST and International Organization for Standardization (ISO)—has helped formalize the principles of fairness, accountability, transparency, explainability, privacy, and human agency. These are now recognized as non-negotiable pillars that must be embedded throughout the AI lifecycle.

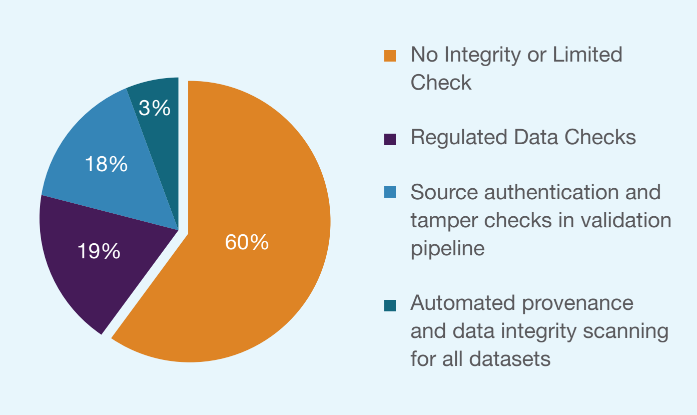
However, while the importance of responsible AI is widely acknowledged, our recent industry survey reveals a significant gap between awareness and implementation. For instance, fewer than 20% of organizations surveyed in India reported that their AI systems offer explainability and transparency or have mechanisms to detect and address bias and fairness concerns. In contrast, over 41% admitted their systems lacked such capabilities—highlighting a troubling absence of foundational safeguards.

The situation is equally concerning in the realm of data privacy and protection. Only 26% of organizations surveyed in India have integrated data masking and personally identifiable information (PII) scanning into their AI development lifecycle, raising critical concerns about regulatory compliance and ethical data stewardship. Furthermore, 60% of respondents indicated they had no formal means to assure the integrity of their AI models, and only 15% had partial mechanisms in place. This lack of model integrity assurance poses risks to reliability, security, and trustworthiness

Data Governance in AI Model Development



Dataset Validation for AI Training



Training data quality and authenticity—the foundation upon which AI performance rests—remain dangerously under-governed. A striking 60% of organizations surveyed in India operate without formal dataset validation processes, relying on informal checks or no verification at all. The result: models vulnerable to degraded performance, embedded bias, and data-integrity failures including tampering.

These findings point to a fundamental requirement: responsible AI demands more than technical excellence. Organizations must cultivate cultures characterized by stewardship, systematic oversight, and deep ethical accountability. This cultural transformation requires governance structures that go beyond algorithmic fixes—frameworks that normalize ethical decision-making, institutionalize continuous monitoring, and embed stakeholder engagement as core operational practice rather than peripheral activity.

When implemented effectively, responsible AI strengthens organizational governance by providing clear mechanisms for oversight, risk management, and regulatory compliance. It promotes fairness by identifying and mitigating bias in datasets and algorithms, ensuring equitable outcomes across diverse populations. Transparency and explainability empower users and regulators to understand how AI decisions are made, fostering trust and enabling informed oversight. Accountability ensures that organizations can assess, correct, and improve AI systems when unintended consequences arise.

Best Practices for Responsible AI

Area	Specific Actionable Steps
Training Data Quality	• **Track data provenance:** Record source and lineage of all datasets. • **Validate automated data:** Use schema checks and anomaly detection tools. • **Conduct third-party assessments:** Engage external assessors to validate dataset authenticity. • **Refresh periodically:** Update datasets to reflect current and unbiased information.
Privacy and Data Protection	• **Automate Data masking and PII-scanning:** Automate detection and masking of personally identifiable information. • **Apply Differential privacy:** Add noise to datasets to prevent re-identification. • **Ensure secure data storage:** Encrypt data using advanced encryption Standard (AES)-256 and enforce strict access controls. • **Conduct compliance checks:** Validate adherence to General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), and other relevant regulations.
Fairness and Bias Mitigation	• **Conduct bias assessments:** Use tools to detect bias in datasets and models. • **Sample diverse data:** Ensure training datasets represent all relevant demographic groups. • **Apply bias correction techniques:** Apply re-weighting or adversarial debiasing during model training. • **Conduct continuous monitoring:** Track fairness metrics post-deployment and retrain models when drift occurs.
Model Integrity and Reliability	• **Use cryptographic hashing:** Use secure hash algorithm (SHA)-256 to verify model artifacts and prevent tampering. • **Apply version control:** Implement tools for tracking model versions. • **Conduct adversarial testing:** Simulate attacks to test model robustness. • **Perform continuous validation:** Monitor performance and retrain models when degradation is detected.

Transparency and Explainability	• **Integrate explainability tools:** Use tools to provide feature-level explanations for model decisions. • **Maintain model documentation:** Maintain detailed model cards and datasheets for datasets. • **Provide user-friendly reports:** Provide clear, non-technical explanations for end-users and regulators. • **Log audit trails:** Log all model decisions for accountability and compliance.
Accountability and Governance	• **Assign AI risk owners:** Designate responsible individuals for each AI system. • **Establish governance committees:** Include cross-functional teams (legal, compliance, ethics, tech). • **Create AI specific incident response playbooks and plan for ethical failures:** Create escalation workflows for bias or harm-related incidents. • **Conduct regular ethical reviews:** Conduct quarterly reviews of AI systems against organizational principles and regulations.

Securing the AI Lifecycle:
From Development to Deployment and Beyond

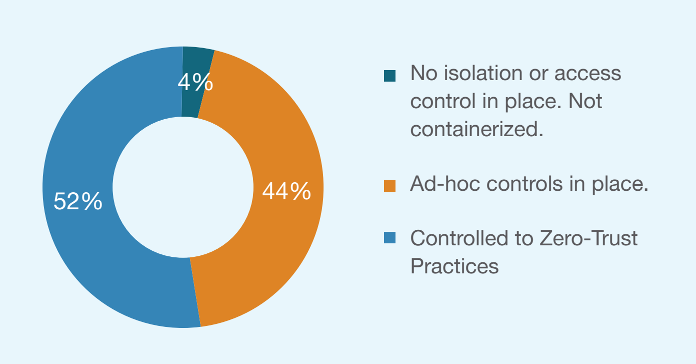
The AI adoption journey has entered a new phase. Organizations in India are moving beyond foundational governance discussions to operationalize responsible AI across complete system lifecycles. While early work on ethics, fairness, and governance established accountability frameworks, production deployment reveals new operational challenges: securing data and models, ensuring deployment integrity, and implementing continuous monitoring and compliance at scale.

Secure Development: Building AI On a Strong Foundation

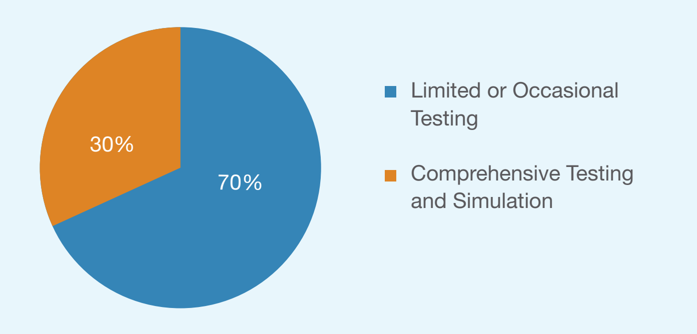
Secure and responsible AI begins in the development environment, where data is curated, models are trained, and foundational decisions are made. This phase is particularly vulnerable to threats such as data poisoning, model tampering, and unauthorized access, all of which can compromise integrity, fairness, and trust.

Encouragingly, 52% of surveyed organizations in India have implemented secured development environments with access controls. Isolation and role-based access prevent unauthorized manipulation of training data or model parameters, while secure authentication and environment segmentation ensure only authorized personnel interact with sensitive components.

Security Controls in AI Development and Deployment Environment



Adoption of Security Testing Practices in Deployed Models



However, advanced practices remain underdeveloped. Fewer than 30% of organizations surveyed in India conduct penetration testing or red teaming to uncover vulnerabilities in data pipelines, model logic, and infrastructure. Data poisoning—where malicious actors inject corrupted data into training sets—is another growing concern, yet only 19% have procedures to detect and mitigate such risks. Organizations should adopt input validation, data provenance tracking, and automated anomaly detection to flag suspicious patterns and maintain data integrity.

Models also require protection. Unauthorized modifications or adversarial interference can lead to harmful outcomes. Techniques such as containerization and sandboxing help isolate model training and experimentation, ensuring separation and reproducibility.

Finally, integrating secure development practices—such as code reviews, dependency scanning, and infrastructure hardening—into the AI lifecycle significantly reduces the attack surface. Combined with a culture of security awareness and ethical responsibility, these measures form the first line of defense for trustworthy AI systems.

Best Practice for Data and Model Development Security

Area	Specific Actionable Steps
Secure Development Environment	• **Conduct environment segmentation:** Separate development, test, and production environments with strict network isolation. • **Implement role-based access control (RBAC):** Assign granular permissions for developers, data scientists, and admins. • **Apply multi-factor authentication (MFA):** Enforce MFA for all development-environment access. • **Encrypt Data:** Use AES-256 encryption for data at rest and Transport Layer Security (TLS) 1.2+ for data in transit.
Secure Development Practices	• **Conduct code reviews:** Enforce peer and tool based - e.g., IAST, SAST, DAST- reviews for all AI-related code changes. • **Perform dependency scanning:** Use tools to detect vulnerable libraries. • **Strengthen infrastructure hardening:** Disable unused ports, enforce least privilege, and apply security patches regularly. • **Deliver security training:** Conduct quarterly workshops on adversarial Machine Learning (ML) threats and secure coding practices.
Data Integrity and Poisoning Prevention	• **Input validation:** Apply schema checks and whitelist rules for incoming data. • **Track data provenance:** Use cryptographic hashing (SHA-256) to verify dataset integrity. • **Deploy anomaly detection:** Deploy ML-based tools to flag outliers or suspicious patterns in training data.
Model Protection	• **Implement containerization:** Use Docker or Kubernetes to isolate model training environments. • **Apply sandboxing:** Restrict external network calls during training to prevent injection attacks. • **Apply version control:** Store model artifacts in tool with immutable commit history. • **Enforce access restrictions:** Limit write permissions to model parameters and scripts to authorized personnel only.
Penetration Testing and Red Teaming	• **Schedule quarterly pen tests:** Include model Application Programming Interface (APIs), data pipelines, and infrastructure. • **Conduct red-team exercises:** Simulate adversarial attacks like model inversion or poisoning. • **Document remediation workflow:** Document vulnerabilities, assign owners, and track fixes in a ticketing system.

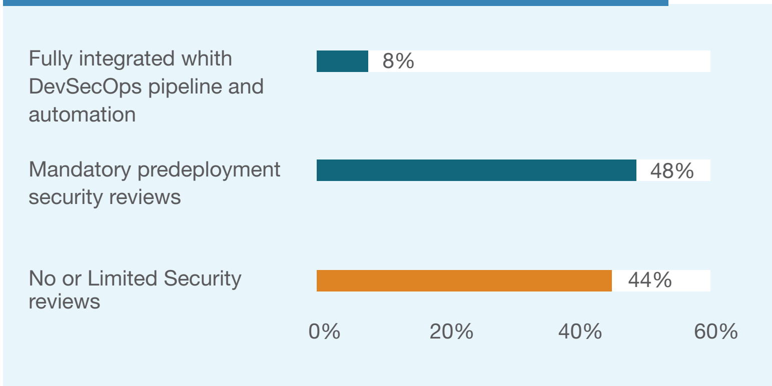
Deployment and Operationalization: Ensuring Safe Transitions

After validating models, organizations must deploy them into production environments where they interact with live data and users. This phase introduces challenges such as version control, change management, and real-time security threats.

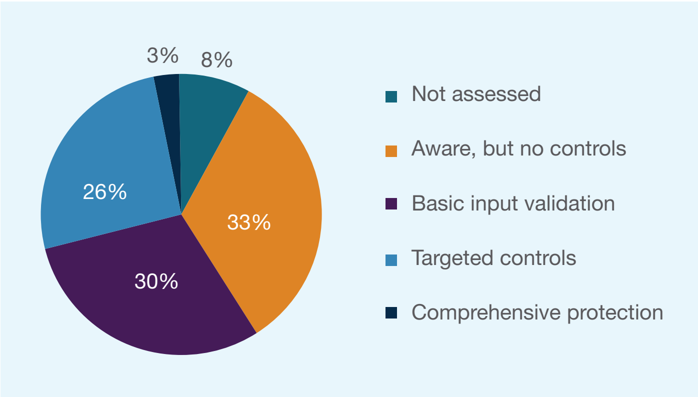
Effective version control and change management are essential to track model evolution, roll back faulty deployments, and maintain consistency. Organizations should maintain detailed logs of model updates, training-data changes, and performance metrics for traceability.

Before deployment, models should undergo rigorous security reviews. While widely recognized as best practice, only 56% of surveyed organizations in India mandate such reviews—manual or automated—before deployment. These reviews typically include vulnerability scans, compliance checks, and validation to ensure models do not expose sensitive data.

Security Review Practices Before AI Model Deployment



Security Measures for Prompt Injection in AI and LLM Deployments



Prompt-injection attacks—especially in generative AI systems—are a growing concern. These attacks manipulate inputs to produce harmful outputs, bypassing safety filters. Mitigation strategies include input sanitization, prompt validation, and layered access controls. Yet only 30% of surveyed organizations have targeted controls, and 59% rely on basic validation measures, indicating limited maturity.

Hallucination risk—where models generate false or misleading outputs—remains a critical concern. Despite its prevalence, only 19% of surveyed organizations have real-time detection or feedback loops in place, while 51% have no controls at all, creating reputational and operational risks.

Exposure of sensitive or regulated data during training or inference must be tightly controlled. Techniques such as data masking, differential privacy, and secure multi-party computation help prevent unauthorized access. However, adoption remains low: only 15% of organizations use automated data controls, while 44% still rely on traditional role-based access—insufficient for modern AI risks.

These gaps underscore the need for advanced data protection and proactive governance in AI operationalization.

Best Practices for Model Deployment and Operationalization

Area	Specific Actionable Steps
Version Control and Change Management	• **Use tools for model registry:** Track versions, metadata, and lineage. • **Enable rollback in continuous integration/continuous deployment (CI/CD) pipelines:** Automate rollbacks to last stable version. • **Maintain immutable logs:** Store logs of model updates, training data changes, and performance metrics in secure storage.
Pre-Deployment Security Reviews	• **Automate security checks:** Integrate vulnerability scanning tools in CI/CD. • **Validate compliance:** Verify adherence to GDPR, HIPAA, or sector-specific standards. • **Test data exposure:** Run tests to ensure models do not leak sensitive data during inference.
Prompt Injection Mitigation	• **Implement input sanitization:** Strip harmful tokens and enforce strict schema validation. • **Apply prompt validation:** Use regex-based filters and allowlists for safe commands. Implement real-time prompt monitoring tool. • **Enforce layered access controls:** Implement RBAC and MFA for prompt submission interfaces. • **Conduct real-time monitoring:** Deploy anomaly detection for suspicious input patterns.
Hallucination Risk Management	• **Display confidence scoring:** Display confidence levels for outputs and flag low-confidence predictions. • **Develop fallback protocols:** Route uncertain outputs to human review or alternative models. • **Implement continuous feedback loops:** Collect user feedback and retrain models periodically. • **Integrate hallucination detection tools:** Integrate Large Language Model (LLM) specific monitoring solutions.
Data Protection	• **Apply data masking and encryption:** Apply AES-256 encryption and anonymization for sensitive data. • **Establish differential privacy:** Add noise to datasets to prevent re-identification. • **Secure multi-party computation:** Use cryptographic techniques for collaborative training. • **Implement automated data governance:** Implement policy-driven access controls beyond RBAC.

Monitoring and Compliance: Sustaining Trust Post-Deployment

AI systems are dynamic by nature. Their performance can drift due to changes in input data, user behaviour, or external conditions. Continuous monitoring is, therefore, essential to keep models accurate, fair, and secure after deployment.

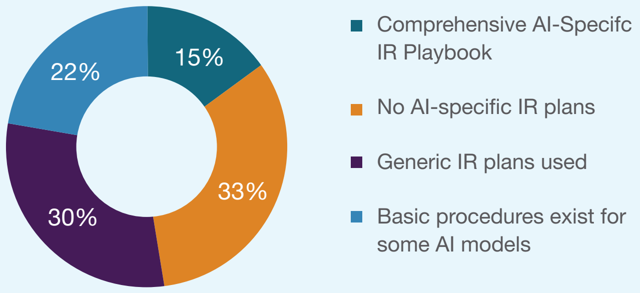
Yet maturity in this area is uneven. 26% of surveyed organizations report no monitoring for security anomalies or performance degradation, and another 45% have only limited or non-real-time monitoring. This lack of oversight increases the risk of undetected failures or malicious activity.

While 40.74% of surveyed organizations in India implement or review governance frameworks, 74% lack real-time monitoring and about 70% have no AI-specific incident-response capabilities. In effect, governance operates without the tools to enforce or validate itself.

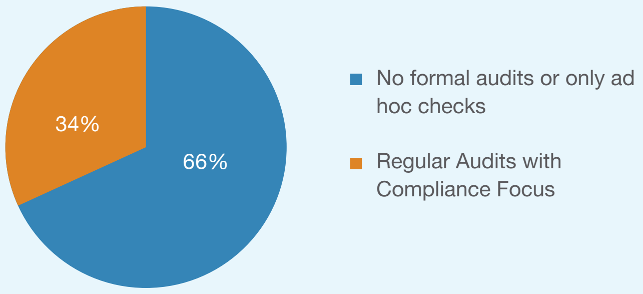
Monitoring must address both technical and ethical dimensions. Technically, organizations should track performance, latency, error rates, and resource usage. Ethically, they must monitor for bias, fairness violations, and unintended consequences. This requires auditable logs, governance dashboards, and feedback loops for real-time insights and corrective actions. However, adoption remains in early stages for many organizations.

A strong incident-response (IR) playbook specific to AI scenarios is equally critical. Whether a model produces biased outputs, leaks sensitive data, or behaves unpredictably, predefined protocols for escalation, investigation, and resolution are essential. Yet only 15% of surveyed organizations in India have AI-specific IR playbook, while 33% have none and 30% rely on generic playbooks. Clear roles, communication strategies, and remediation workflows are critical for swift, transparent action.

Adoption of AI-Specific Incident Response (IR) Playbooks



AI System Auditing Practices for Regulatory Compliance



Periodic assessments are another cornerstone of AI compliance. These assessments should evaluate performance, data integrity, governance, and regulatory alignment. Internal reviews must be complemented by third-party assessments for independent validation. Alarmingly, 66% of surveyed organizations conduct no formal audits or only ad hoc checks, undermining compliance and exposing them to regulatory and ethical risks.

As global regulations such as the EU AI Act, OECD guidelines, and NIST standards evolve, compliance will become increasingly complex. Organizations must embed governance and documentation into every phase of the AI lifecycle and foster a culture of ethical accountability. Current gaps in monitoring, incident response, and auditing highlight the urgent need to shift from reactive to proactive governance to ensure trust, resilience, and regulatory alignment.

Best Practices for Monitoring, Incident Reporting, and Compliance

Area	Specific Actionable Steps
Continuous Monitoring	• **Deploy real-time monitoring tools:** Track performance, latency, and error rates. • **Implement ethical monitoring:** Detect bias and fairness violations in model outputs. • **Maintain auditable logs:** Record model events and decisions to ensure transparency and regulatory compliance. • **Incorporate user feedback loops:** Capture end-user reports to surface unexpected behavior and usability issues. • **Monitor for privacy violations:** Detect leakage or disclosure risks in generative and inference outputs.
Incident Response (IR)	• **Develop AI IR playbooks:** Define escalation paths and remediation steps for AI incidents. • **Integrate threat intelligence feeds:** Consume sources (e.g., CERT-In, MITRE) to track emerging AI threats. • **Define IR roles and responsibilities:** Assign ownership for detection, triage, escalation, and remediation. • **Establish communication protocols:** Specify internal and external notification channels and messaging templates. • **Conduct IR drills:** Run regular simulations to test readiness and refine response workflows. • **Specify regulator-notification procedures:** Define thresholds, timelines and reporting channels for major incidents (e.g., GDPR, EU AI Act).
Compliance and Assessments	• **Schedule recurring AI assessments:** Conduct quarterly or semi-annual reviews of performance and governance. • **Engage third-party assessors:** Obtain independent validation of model performance, fairness, and security. • **Maintain comprehensive documentation:** Record model specs, data provenance, and compliance certifications. • **Update policies continuously:** Align governance with evolving standards and regulations (e.g., EU AI Act, NIST, ISO). • **Incorporate fairness assessments:** Embed fairness testing into regular compliance reviews.
Governance Culture	• **Implement governance dashboards:** Provide leadership with real-time visibility into model risk and performance. • **Train teams on AI risk and ethics:** Deliver role-based training on risk management and ethical use. • **Create monitoring-to-improvement feedback loops:** Turn monitoring insights into prioritized remediation actions. • **Establish a cross-functional ethics committee:** Review and approve high-risk AI use cases and mitigation plans. • **Integrate responsible AI KPIs:** Embed governance metrics into performance reviews and team objectives.

AI has moved from the horizon to here and now—actively transforming industry dynamics, reinventing operational models, and creating new innovation trajectories. This momentum, however, demands parallel accountability for the risk landscape that expands proportionally with the scale of AI. Technical vulnerabilities become more exploitable, ethical considerations more complex, and regulatory expectations more demanding.

Successful AI adoption is not only about building powerful models or deploying cutting-edge algorithms. It is equally about embedding trust, transparency, and accountability across the entire AI lifecycle—from secure development and responsible deployment to continuous monitoring and governance.

Organizations in India that proactively invest in robust governance frameworks, secure development practices, and responsible AI principles will not only mitigate risks but also gain a strategic edge in a rapidly evolving digital landscape. The path forward demands proactive stewardship, ensuring that AI systems are not just intelligent but ethical, resilient, and aligned with societal values.

Notes

OUR LEADERS

DHRUV PHOPHALIA
Managing Director and Leader -
Disputes and Investigations
+91-9820460731
dphophalia@alvarezandmarsal.com

CHANDRA PRAKASH SURYAWANSHI
Managing Director and India Global
Cyber Risk Services Practice Leader
+91-9900020190
csuryawanshi@alvarezandmarsal.com

RAHUL GOSAIN
Managing Director,
Disputes and Investigations
+91-9811522554
rgosain@alvarezandmarsal.com

ABOUT ALVAREZ & MARSAL

Founded in 1983, Alvarez & Marsal is a leading global professional services firm. Renowned for its leadership, action and results, Alvarez & Marsal provides advisory, business performance improvement and turnaround management services, delivering practical solutions to address clients' unique challenges. With a world-wide network of experienced operators, world-class consultants, former regulators and industry authorities, Alvarez & Marsal helps corporates, boards, private equity firms, law firms and government agencies drive transformation, mitigate risk and unlock value at every stage of growth.

Follow A&M on:

To learn more, visit: **AlvarezandMarsal.com**

ALVAREZ & MARSAL
LEADERSHIP. **ACTION. RESULTS.**℠