AI GOVERNANCE AUDIT CHECKLIST



By: Dr. Nath Alagbe

AI Governance Audit

AI Governance Audit is an independent and objective assessment of an organization's AI governance framework. It evaluates the design and operating effectiveness of controls intended to manage AI-related risks, ensure compliance, and promote responsible AI practices. The audit aims to provide assurance to stakeholders that AI systems are developed and used in a controlled and trustworthy environment.

AI Governance Audit Checklist

This section presents an abridged AI Governance Audit Checklist, organized by domain.

Author: Dr. Nath Alagbe; CISA, CISM, CISSP, CRISC, CCAK, AAIA, CCEP, CFE, FCA, FCTI, FFAR, MBA, PhD

AI Governance Structure and Oversight

Control Area	Risk Description	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
Board Oversight of AI Strategy	Lack of board understanding or engagement in AI strategy leading to misaligned AI initiatives and unmanaged risks.	To ensure the Board of Directors actively oversees and approves the organization's AI strategy and risk appetite.	1. Review board meeting minutes for discussions and approvals related to AI strategy and risk. 2. Interview board members or relevant executives on their understanding and involvement in AI governance.	Board meeting minutes, AI strategy documents, executive presentations on AI, and interview transcripts.	NIST AI RMF (Govern), ISO/IEC 42001 (4.1, 5.1)
AI Governance Committee Establishment	Absence of a dedicated committee to oversee AI governance, leading to fragmented decision-making and inconsistent policy application.	To ensure a formal AI Governance Committee is established with a clear mandate, roles, and responsibilities.	1. Review the charter and meeting minutes of the AI Governance Committee. 2. Interview committee members to confirm understanding of their roles and responsibilities.	Committee charter, meeting minutes, attendance records, and interview transcripts.	NIST AI RMF (Govern), ISO/IEC 42001 (5.2)
Defined Roles and Responsibilities	Unclear or overlapping roles and responsibilities for AI governance, leading to accountability gaps and inefficient operations.	To ensure clear definition and communication of roles, responsibilities, and accountability for AI governance across the organization.	1. Review organizational charts, job descriptions, and RACI matrices related to AI governance. 2. Interview key personnel to confirm their understanding of their AI governance roles.	Organizational charts, job descriptions, RACI matrices, internal communications, and interview transcripts.	NIST AI RMF (Govern), ISO/IEC 42001 (5.3)
Integration with Enterprise Governance	AI governance operating in isolation from broader enterprise governance frameworks, leading to inconsistencies and missed synergies.	To ensure AI governance is integrated with existing enterprise governance, risk, and compliance (GRC) frameworks.	1. Review documentation outlining the integration points between AI governance and enterprise GRC. 2. Interview GRC and AI governance leads to confirm integration practices.	GRC framework documentation, AI governance framework, integration policies, and interview transcripts.	COBIT (Align, Plan, and Organize), COSO (Control Environment)

Resource Allocation for AI Governance	Insufficient allocation of human, financial, and technological resources for effective AI governance.	To ensure adequate resources are allocated to support AI governance functions, including staffing, training, and tools.	1. Review budget allocations for AI governance initiatives. 2. Interview AI governance leads on resource adequacy and training needs.	Budget reports, training plans, staffing reports, tool procurement records, and interview transcripts.	NIST AI RMF (Govern), ISO/IEC 42001 (7.1)
AI Ethics Committee Effectiveness	An AI Ethics Committee that is ineffective or lacks influence, leading to unaddressed ethical concerns in AI systems.	To ensure the AI Ethics Committee operates effectively, providing meaningful guidance and oversight on ethical AI development and use.	1. Review AI Ethics Committee charter, meeting minutes, and records of ethical impact assessments. 2. Interview committee members and AI project teams on the committee's impact.	Committee charter, meeting minutes, ethical impact assessment reports, project documentation, and interview transcripts.	IEEE AI Ethics, OECD AI Principles

AI Policies and Standards

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
AI Ethics Policy	Absence or inadequacy of an AI ethics policy, leading to unethical AI development and deployment.	To ensure a comprehensive AI ethics policy is established, approved, and communicated, guiding responsible AI behavior.	1. Review the AI Ethics Policy for completeness and alignment with ethical principles. 2. Interview AI development teams on their understanding and application of the policy.	AI Ethics Policy document, approval records, training materials, and interview transcripts.	IEEE AI Ethics, OECD AI Principles
Data Governance Policy for AI	Insufficient data governance policies tailored for AI, leading to data quality issues, privacy breaches, and biased models.	To ensure robust data governance policies are in place for AI, covering data quality, privacy, security, and ethical use.	1. Review data governance policies specifically for AI data. 2. Examine data quality reports and privacy impact assessments for AI systems.	Data governance policies, data quality reports, PIA reports, and data flow diagrams.	GDPR, CCPA, ISO/IEC 42001 (8.2)
AI Development Standards	Lack of standardized development practices for AI, leading to inconsistent quality, security vulnerabilities, and difficulties in maintenance.	To ensure standardized development practices and guidelines are established and followed for all AI systems.	1. Review AI development guidelines, coding standards, and model documentation templates. 2. Interview AI engineers on their adherence to established standards.	Development guidelines, code repositories, model documentation, peer review records, and interview transcripts.	NIST AI RMF (Map), ISO/IEC 42001 (8.1)
Acceptable Use Policy for AI	Absence of a clear acceptable use policy for AI, leading to misuse of AI systems and potential legal or ethical breaches.	To ensure an acceptable use policy for AI is established and communicated, preventing misuse and promoting responsible application.	1. Review the Acceptable Use Policy for AI. 2. Conduct awareness sessions and assess employee understanding of the policy.	Acceptable Use Policy, training materials, employee acknowledgments, and incident reports.	NIST AI RMF (Govern), ISO/IEC 42001 (5.2)
Policy Communication and Training	Ineffective communication and training on AI policies, leading to	To ensure all relevant personnel are aware of and trained on AI	1. Review training records and communication logs for AI policies.	Training attendance records, communication	ISO/IEC 42001 (7.3)

	low awareness and non-compliance among employees.	governance policies and standards.	2. Conduct surveys or quizzes to assess employee knowledge of AI policies.	logs, policy awareness surveys, and quiz results.	
Policy Review and Update Process	Outdated or irrelevant AI policies due to lack of a systematic review and update process.	To ensure AI policies and standards are regularly reviewed, updated, and approved to remain relevant and effective.	1. Review the policy review schedule and records of policy updates. 2. Interview policy owners on their process for monitoring regulatory changes and emerging risks.	Policy review schedule, version control records, change logs, regulatory watch reports.	ISO/IEC 42001 (10.2)

AI Risk Management

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
AI Risk Identification Process	Failure to systematically identify all relevant AI-specific risks, leading to unmanaged exposures.	To ensure a comprehensive and systematic process for identifying AI-specific risks across the organization.	1. Review the AI risk identification methodology and tools used. 2. Interview AI project teams and risk managers on their risk identification practices.	Risk identification methodology, workshop outputs, brainstorming sessions, and interview transcripts.	NIST AI RMF (Map), ISO/IEC 42001 (6.1)
AI Risk Assessment Methodology	Inadequate methodology for assessing AI risks, leading to inaccurate risk ratings and mis-prioritized mitigation efforts.	To ensure a robust and consistent methodology is used for assessing the likelihood and impact of AI risks.	1. Review the AI risk assessment methodology, including criteria for likelihood and impact. 2. Examine sample AI risk assessments for adherence to the methodology.	Risk assessment methodology document, risk assessment reports, risk scoring matrices.	NIST AI RMF (Measure), ISO/IEC 42001 (6.1)
AI Risk Register Management	Poor management of the AI risk register, leading to outdated information, unassigned ownership, and ineffective tracking.	To ensure a centralized, up-to-date, and actively managed AI risk register with clear ownership and mitigation plans.	1. Review the AI risk register for completeness, accuracy, and currency. 2. Interview risk owners on their management of assigned AI risks.	AI risk register, risk ownership assignments, mitigation plans, status updates.	NIST AI RMF (Manage), ISO/IEC 42001 (6.1)
AI Risk Mitigation Strategies	Ineffective or unimplemented strategies for mitigating identified AI risks.	To ensure appropriate and effective mitigation strategies are developed and implemented for all identified AI risks.	1. Review AI risk mitigation plans and evidence of their implementation. 2. Interview risk owners on the effectiveness of their mitigation actions.	Mitigation plans, project documentation, control implementation evidence, and effectiveness reports.	NIST AI RMF (Manage), ISO/IEC 42001 (6.1)
Emerging AI Risk Monitoring	Failure to monitor and assess emerging AI risks, leading to unforeseen exposures and lack of preparedness.	To ensure a proactive process for monitoring and assessing emerging AI risks and their potential impact on the organization.	1. Review threat intelligence reports, industry publications, and research on emerging AI risks. 2. Interview risk managers on their process for incorporating emerging risks into the risk register.	Threat intelligence reports, research papers, emerging risk watchlists, risk register updates.	NIST AI RMF (Govern), ISO/IEC 42001 (6.1)

AI Regulatory Compliance

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
Regulatory Landscape Monitoring	Failure to monitor and adapt to the evolving AI regulatory landscape, leading to non-compliance.	To ensure a systematic process for monitoring, analyzing, and responding to changes in AI-related laws and regulations.	1. Review regulatory intelligence reports and legal updates related to AI. 2. Interview legal and compliance teams on their process for tracking regulatory changes.	Regulatory watch reports, legal advisories, compliance bulletins, meeting minutes.	EU AI Act, GDPR, CCPA
Compliance Impact Assessments	Lack of formal assessments to determine the impact of new AI regulations on existing AI systems and processes.	To ensure that formal compliance impact assessments are conducted for new or updated AI regulations.	1. Review compliance impact assessment reports for key AI regulations. 2. Interview AI project teams on how regulatory requirements are incorporated into their work.	Compliance impact assessment reports, project plans, requirements traceability matrices.	EU AI Act, GDPR, CCPA
AI Compliance Controls Implementation	Failure to implement adequate controls to meet AI regulatory requirements.	To ensure that effective controls are designed and implemented to achieve compliance with applicable AI laws and regulations.	1. Review control documentation and evidence of control implementation for key regulatory requirements. 2. Test the operating effectiveness of selected AI compliance controls.	Control matrices, policy documents, system configurations, audit logs, test results.	EU AI Act, GDPR, CCPA
Regulatory Reporting	Inaccurate or untimely reporting to regulatory authorities regarding AI systems.	To ensure accurate and timely reporting to regulatory authorities as required by AI laws and regulations.	1. Review regulatory submissions and correspondence related to AI. 2. Interview compliance officers on their regulatory reporting processes for AI.	Regulatory reports, submission confirmations, and correspondence with regulators.	EU AI Act, GDPR, CCPA
Legal Counsel Involvement	Insufficient involvement of legal counsel in AI development and deployment, leading to unmanaged legal risks.	To ensure legal counsel is appropriately involved throughout the AI lifecycle to advise on compliance and legal risks.	1. Review records of legal reviews for AI projects and contracts. 2. Interview AI project leads and legal counsel on their collaboration.	Legal review reports, contract reviews, meeting minutes with legal, interview transcripts.	All relevant AI regulations
Data Protection Officer (DPO) Engagement	Inadequate engagement of the Data Protection Officer (DPO) in AI systems processing personal data.	To ensure the DPO is actively involved in the design, development, and deployment of AI systems that process personal data.	1. Review records of DPO consultations for AI projects. 2. Interview the DPO and AI project teams on their collaboration.	DPO consultation logs, PIA reports, meeting minutes with DPO.	GDPR, CCPA

AI Ethics and Responsible AI

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
Ethical Impact Assessments (EIAs)	Failure to conduct EIAs for AI systems, leading to unforeseen ethical	To ensure that Ethical Impact Assessments (EIAs) are systematically	1. Review EIA reports for AI projects. 2. Interview AI project teams on their process for	EIA reports, ethical risk registers, project documentation,	IEEE AI Ethics, OECD AI Principles

	harms and reputational damage.	conducted for all AI systems, especially high-risk ones.	identifying and mitigating ethical risks.	stakeholder consultation records.	
Fairness and Bias Detection	Inadequate mechanisms for detecting and measuring bias in AI models and data, leading to discriminatory outcomes.	To ensure robust mechanisms are in place for detecting, measuring, and reporting algorithmic bias and fairness metrics.	1. Review bias detection tools and reports for AI models. 2. Examine fairness metrics and their thresholds.	Bias audit reports, fairness metrics dashboards, data quality reports, and model validation reports.	NIST AI RMF (Measure), ISO/IEC 42001 (8.2)
Bias Mitigation Strategies	Failure to implement effective strategies to mitigate identified biases in AI systems.	To ensure that appropriate and effective strategies are developed and implemented to mitigate identified algorithmic biases.	1. Review bias mitigation plans and evidence of their implementation. 2. Interview AI development teams on the effectiveness of mitigation actions.	Bias mitigation plans, model retraining logs, data re-sampling documentation, and effectiveness reports.	NIST AI RMF (Manage), ISO/IEC 42001 (8.2)
Transparency and Explainability	Lack of transparency and explainability in AI systems, leading to distrust and inability to understand AI decisions.	To ensure AI systems are designed and operated with appropriate levels of transparency and explainability.	1. Review XAI (Explainable AI) documentation and tools used. 2. Interview end-users on their understanding of AI decisions.	XAI reports, model documentation, user feedback, and explainability tool outputs.	NIST AI RMF (Map), ISO/IEC 42001 (8.3)
Human Oversight Mechanisms	Insufficient human oversight mechanisms for AI systems, leading to autonomous operation without adequate human control.	To ensure meaningful human oversight is maintained for AI systems, with clear intervention and override capabilities.	1. Review documentation of human-in-the-loop/on-the-loop processes. 2. Observe human operators interacting with AI systems and their intervention capabilities.	Process documentation, training materials, incident logs (for human intervention), and observation reports.	NIST AI RMF (Govern), ISO/IEC 42001 (8.3)
Accountability Framework for AI	Unclear accountability for AI system outcomes, leading to diffused responsibility and lack of ownership for errors.	To ensure a clear accountability framework is established for AI systems, assigning responsibility for their outcomes.	1. Review accountability matrices and incident reports for AI systems. 2. Interview AI project leads and business owners on their accountability for AI outcomes.	Accountability matrices, incident reports, project charters, and interview transcripts.	NIST AI RMF (Govern), ISO/IEC 42001 (5.3)

AI Inventory and Asset Management

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
Comprehensive AI Inventory	Incomplete or inaccurate inventory of AI systems, leading to shadow AI and unmanaged risks.	To ensure a comprehensive and up-to-date inventory of all AI systems and their components within the organization.	1. Review the AI asset register and compare it against deployed AI systems. 2. Conduct interviews with business units to identify any unrecorded AI initiatives.	AI asset register, deployment logs, business unit surveys, interview transcripts.	NIST AI RMF (Map), ISO/IEC 42001 (8.1)
AI System Classification	Inconsistent or inappropriate classification of AI systems, hindering	To ensure AI systems are consistently classified based on risk, criticality, and	1. Review the AI system classification framework and sample classifications.	Classification framework, AI system inventory with	NIST AI RMF (Map), ISO/IEC 42001 (8.1)

	risk prioritization and resource allocation.	other relevant attributes.	2. Interview AI project leads on their understanding and application of the classification scheme.	classifications, training materials.	
AI System Documentation	Inadequate documentation for AI systems, leading to difficulties in understanding, maintaining, and auditing them.	To ensure comprehensive and standardized documentation is maintained for each AI system throughout its lifecycle.	1. Review AI system documentation (design, data, models, deployment). 2. Interview AI engineers on their documentation practices.	Design documents, data dictionaries, model cards, deployment guides, and user manuals.	NIST AI RMF (Map), ISO/IEC 42001 (8.1)
Data Lineage for AI	Lack of clear data lineage for AI training and operational data, hindering data quality and bias investigations.	To ensure clear data lineage is established and documented for all data used by AI systems.	1. Review data lineage documentation for key AI models. 2. Interview data engineers on their processes for tracking data sources and transformations.	Data lineage maps, data flow diagrams, and data transformation scripts.	GDPR, CCPA, ISO/IEC 42001 (8.2)
AI Asset Ownership	Unclear ownership of AI assets, leading to accountability gaps and unmanaged risks.	To ensure clear ownership is assigned for all AI assets, including models, data, and infrastructure.	1. Review AI asset ownership records. 2. Interview AI project leads and business owners on their understanding of asset ownership.	AI asset register with owners, project charters, and RACI matrices.	NIST AI RMF (Govern), ISO/IEC 42001 (5.3)

AI Lifecycle Governance

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
Design Phase Governance	Lack of governance considerations in the AI design phase, leading to ethical, security, and compliance issues downstream.	To ensure ethical, security, and compliance considerations are embedded early in the AI system design phase.	1. Review design documents and requirements for AI projects for governance considerations. 2. Interview AI architects and designers on their process for incorporating governance.	Design documents, requirements specifications, architectural diagrams, and ethical review records.	NIST AI RMF (Govern, Map), ISO/IEC 42001 (8.1)
Data Collection and Preparation	Inadequate controls over data collection and preparation, leading to biased, inaccurate, or non-compliant AI models.	To ensure data used for AI is collected, prepared, and managed in a secure, compliant, and ethical manner.	1. Review data collection procedures, consent mechanisms, and data quality checks. 2. Examine data lineage documentation and anonymization techniques.	Data collection procedures, consent logs, data quality reports, data lineage maps, and anonymization scripts.	GDPR, CCPA, ISO/IEC 42001 (8.2)
Model Development and Training	Lack of standardized practices and controls during model development and training, leading to unreliable or insecure models.	To ensure AI models are developed and trained using standardized, secure, and documented practices.	1. Review model development guidelines, coding standards, and training logs. 2. Examine version control records for model code and training data.	Development guidelines, training logs, version control repositories, and model documentation.	NIST AI RMF (Map), ISO/IEC 42001 (8.1)
Model Testing and Validation	Insufficient testing and validation of AI models, leading	To ensure AI models undergo rigorous testing and	1. Review model testing plans, test cases, and validation reports. 2.	Test plans, test results, validation reports, bias	NIST AI RMF (Measure),

	to deployment of models that fail to meet performance, fairness, or security requirements.	independent validation before deployment.	Interview independent validators on their methodology and findings.	testing results, and security testing reports.	ISO/IEC 42001 (8.1)
Deployment and Release Management	Uncontrolled deployment of AI models, leading to operational disruptions, security vulnerabilities, and compliance issues.	To ensure AI models are deployed in a controlled, secure, and documented manner, following formal release management processes.	1. Review deployment checklists, approval workflows, and release notes. 2. Examine audit logs for deployment activities.	Deployment checklists, approval records, release notes, and deployment logs.	NIST AI RMF (Manage), ISO/IEC 42001 (8.1)
Continuous Monitoring and Maintenance	Inadequate monitoring of deployed AI models, leading to undetected model drift, performance degradation, and emerging risks.	To ensure deployed AI models are continuously monitored for performance, bias, security, and compliance, and maintained appropriately.	1. Review monitoring dashboards, alert configurations, and maintenance logs. 2. Interview AI operations teams on their monitoring and incident response procedures.	Monitoring dashboards, alert logs, maintenance records, and incident reports.	NIST AI RMF (Measure, Manage), ISO/IEC 42001 (9.1)

AI Data Governance

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
Data Quality Management	Poor data quality used for AI training and operations, leading to inaccurate, unreliable, and biased AI models.	To ensure data used for AI systems meets defined quality standards for accuracy, completeness, consistency, and timeliness.	1. Review data quality policies, standards, and metrics for AI data. 2. Examine data quality reports and remediation logs.	Data quality policies, data profiling reports, data cleansing scripts, and data quality dashboards.	ISO/IEC 42001 (8.2), DAMA DMBOK
Data Privacy and Consent	Processing personal data for AI without proper consent or legal basis, leading to privacy violations and regulatory penalties.	To ensure personal data used in AI systems is processed lawfully, fairly, and transparently, with appropriate consent mechanisms.	1. Review privacy policies, consent forms, and records of processing activities (RoPA) for AI systems. 2. Examine mechanisms for managing data subject rights (e.g., right to be forgotten).	Privacy policies, consent logs, RoPA, data subject request logs, PIA reports.	GDPR, CCPA, ISO/IEC 42001 (8.2)
Data Anonymization and Pseudonymization	Inadequate anonymization or pseudonymization of data used for AI, leading to re-identification risks and privacy breaches.	To ensure appropriate techniques are used to de-identify personal data before it is used for AI training or testing.	1. Review data anonymization policies and procedures. 2. Examine the techniques used (e.g., k-anonymity, differential privacy) and test their effectiveness against re-identification.	Anonymization policies, data masking scripts, and re-identification risk assessment reports.	GDPR, CCPA, ISO/IEC 42001 (8.2)
Data Security and Access Controls	Unauthorized access to AI training data or operational data, leading to data breaches, poisoning, or intellectual property theft.	To ensure robust security controls and access management are in place to protect AI data from unauthorized access, modification, or disclosure.	1. Review data security policies, access control matrices, and encryption standards for AI data. 2. Examine access logs and conduct access reviews for AI data repositories.	Security policies, access control lists, encryption configurations, and access review logs.	ISO 27001, NIST AI RMF (Map)

Data Lineage and Provenance	Lack of visibility into the origin, transformations, and movement of AI data, hindering troubleshooting, audits, and bias investigations.	To ensure clear data lineage and provenance are documented and maintained for all data used in AI systems.	1. Review data lineage documentation and tools used to track data flow. 2. Trace the lineage of a sample dataset from its source to its use in an AI model.	Data lineage maps, data dictionaries, ETL scripts, and data catalog entries.	ISO/IEC 42001 (8.2), DAMA DMBOK
Data Retention and Disposal	Retaining AI data longer than necessary or failing to dispose of it securely, leading to increased privacy risks and storage costs.	To ensure AI data is retained only as long as necessary and disposed of securely in accordance with legal and business requirements.	1. Review data retention policies and schedules for AI data. 2. Examine records of secure data disposal and archival.	Data retention policies, disposal logs, and secure destruction certificates.	GDPR, CCPA, ISO/IEC 42001 (8.2)
Data Bias and Representativeness	Using biased or unrepresentative data for AI training, leading to discriminatory or inaccurate AI models.	To ensure AI training data is evaluated for bias and representativeness, and appropriate mitigation strategies are applied.	1. Review procedures for assessing data bias and representativeness. 2. Examine data profiling reports and bias mitigation actions taken during data preparation.	Data bias assessment reports, data profiling results, and data re-sampling documentation.	NIST AI RMF (Map), ISO/IEC 42001 (8.2)

AI Security Governance

Control Area	Risk Addressed	Control Objective	Audit Test Steps	Evidence Required	Framework Mapping
AI Security Policy and Standards	Absence of specific security policies and standards for AI systems, leading to inconsistent and inadequate security controls.	To ensure comprehensive security policies and standards are established, specifically addressing the unique threats to AI systems.	1. Review AI security policies, standards, and guidelines. 2. Assess their alignment with industry best practices (e.g., OWASP Top 10 for LLMs, MITRE ATLAS).	AI security policy documents, secure coding standards for AI, and alignment mapping documents.	ISO 27001, OWASP Top 10 for LLMs
AI Threat Modeling	Failure to conduct threat modeling for AI systems, leading to unidentified vulnerabilities and inadequate security design.	To ensure threat modeling is systematically performed during the design phase of AI systems to identify and mitigate potential attacks.	1. Review threat modeling methodologies and reports for AI projects. 2. Interview security architects on their process for identifying AI-specific threats.	Threat modeling reports, architectural diagrams, risk mitigation plans.	MITRE ATLAS, NIST AI RMF (Map)
Adversarial Testing and Red Teaming	Lack of adversarial testing and red teaming, leaving AI systems vulnerable to sophisticated attacks (e.g., evasion, poisoning).	To ensure AI systems undergo rigorous adversarial testing and red teaming to identify and remediate vulnerabilities before deployment.	1. Review adversarial testing plans, red teaming reports, and remediation logs. 2. Assess the scope and frequency of these testing activities.	Red teaming reports, penetration testing results, vulnerability remediation logs.	MITRE ATLAS, NIST AI RMF (Measure)
Model Security and Integrity	Inadequate protection of AI models from unauthorized access, modification, or theft, leading to intellectual property loss or compromised system integrity.	To ensure AI models are protected against unauthorized access, tampering, and theft throughout their lifecycle.	1. Review access controls, encryption, and integrity monitoring mechanisms for AI models. 2. Examine audit logs for model access and modifications.	Access control lists, encryption configurations, file integrity monitoring logs, and model version control records.	ISO 27001, NIST AI RMF (Manage)

— Dr. Nath Alagbe

Data Poisoning Prevention	Vulnerability to data poisoning attacks during model training or fine-tuning, leading to compromised model behavior.	To ensure controls are in place to detect and prevent data poisoning attacks, ensuring the integrity of training data.	1. Review data validation procedures, anomaly detection mechanisms, and data provenance tracking. 2. Examine reports on data quality and anomaly detection during training.	Data validation scripts, anomaly detection logs, data provenance records, threat modeling reports.	MITRE ATLAS, NIST AI RMF (Manage)
Prompt Injection and Jailbreak Defenses (for LLMs)	Susceptibility of LLMs to prompt injection and jailbreak attacks, leading to unauthorized actions or disclosure of sensitive information.	To ensure robust defenses are implemented to detect and mitigate prompt injection and jailbreak attempts against LLMs.	1. Review input validation, sanitization, and output filtering mechanisms for LLMs. 2. Examine testing reports specifically targeting prompt injection vulnerabilities.	Input validation rules, output filtering configurations, adversarial testing reports, and incident logs.	OWASP Top 10 for LLMs

Follow me on:

📌 Click 👉 ***here*** for my LinkedIn Profile.

📌 Click 👉 ***here*** for CyberVerge's LinkedIn Page

📌 Click 👉 ***here*** & 👉 ***here*** for X Accounts

— Dr. Nath Alagbe —