Deep Learning Governance Audit Checklist

Type: Audit Template / Checklist

Target Audience: AI Governance Professionals, Risk Managers, Compliance Officers

Audit Information

Field	Details
Organization Name	
AI System Name/ID	
Model Type	
Audit Date	
Lead Auditor	
System Owner	
Audit Scope	☐ Pre-Deployment ☐ Annual Review ☐ Post-Incident ☐ Vendor Assessment

1. Executive Summary

Purpose

This audit checklist provides a structured framework for evaluating deep learning (DL) systems against organizational governance standards, regulatory requirements (such as the EU AI Act and NIST AI RMF), and best practices for responsible AI. Due to the "black box" nature of deep learning, this audit prioritizes documentation, reproducibility, and risk mitigation over pure interpretability.

When to Use This Checklist

- **Gate Review:** Before moving a deep learning model from development to production (Launch Readiness).
- **Periodic Audit:** Annual or quarterly reviews of live systems to detect drift or governance decay.
- **Incident Response:** To identify root causes following a performance failure or bias incident.
- **Procurement:** To assess third-party deep learning vendors.

Scoring & Interpretation

Each item should be marked as **Compliant** (fully met), **Partial** (gaps exist), **Non-Compliant** (not met), or **N/A** (not applicable).

- **Critical Findings:** Any "Non-Compliant" mark in sections marked "Critical" (e.g., Bias Testing, Safety) typically necessitates a No-Go decision for deployment.
- **Scoring:** Calculate the percentage of "Compliant" items against total applicable items. See Section 6 for thresholds.

2. Pre-Deployment Audit

Focus: Ensuring the fundamental documentation and risk assessments are complete before the model enters production.

2.1 Model Documentation

Audit Item	Status	Evidence / Notes
Model Card: A standardized Model Card (or equivalent) exists, detailing intended use, limitations, and model architecture.	☐ Comp ☐ Part ☐ Non ☐ N/A	
System Card: Broader system documentation exists, describing how the model integrates with downstream applications and user interfaces.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Decision Logs: Key design decisions (e.g., choice of architecture, trade-offs between accuracy and speed) are documented.	☐ Comp ☐ Part ☐ Non ☐ N/A	

2.2 Data Governance

Audit Item	Status	Evidence / Notes
Data Lineage: Training data sources are fully documented, including provenance, collection dates, and versioning.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Data Rights: Legal basis for using training data (consent, contract, or legitimate interest) is verified.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Data Quality Assessment: Training data has been assessed for completeness, accuracy, and representativeness.	☐ Comp ☐ Part ☐ Non ☐ N/A	

2.3 Risk & Fairness

Audit Item	Status	Evidence / Notes
Bias Testing: Quantitative testing for disparate impact/bias across protected groups has been performed.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Risk Classification: The system has been classified by risk level (e.g., High, Medium, Low) according to relevant frameworks (e.g., EU AI Act, internal policy).	☐ Comp ☐ Part ☐ Non ☐ N/A	

Failure Modes Analysis: A "pre-mortem" or failure mode analysis has identified how the model might fail and the consequences of such failure.	☐ Comp ☐ Part ☐ Non ☐ N/A	

2.4 Human Oversight

Audit Item	Status	Evidence / Notes
Oversight Model: The level of human oversight (Human-in-the-loop, on-the-loop, or out-of-the-loop) is defined and documented.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Operator Training: Human reviewers have been trained on the system's capabilities, limitations, and automation bias risks.	☐ Comp ☐ Part ☐ Non ☐ N/A	

3. Technical Governance Checklist

Focus: Evaluating the engineering rigor, reproducibility, and stability of the deep learning architecture.

3.1 Architecture & Reproducibility

Audit Item	Status	Evidence / Notes
Architecture Definition: The specific neural network architecture (e.g., Transformer, CNN, LSTM) is documented and justified for the use case.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Code Versioning: All model code is stored in version control (e.g., Git) with a clear commit history linked to the deployed artifact.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Artifact Management: The trained model artifact (weights/parameters) is hashed, versioned, and stored securely (e.g., MLflow, Artifactory).	☐ Comp ☐ Part ☐ Non ☐ N/A	
Reproducibility: A separate team or auditor can reproduce the training run using documented seeds, data versions, and code.	☐ Comp ☐ Part ☐ Non ☐ N/A	

3.2 Hyperparameters & Training

Audit Item	Status	Evidence / Notes
Hyperparameter Logging: All key hyperparameters (learning rate, batch size, epochs, dropout, etc.) are logged.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Train/Test Split: Data leakage prevention is verified; test sets are strictly held out from the training process.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Overfitting Checks: Validation loss curves and metrics demonstrate that the model is not overfitted to training data.	☐ Comp ☐ Part ☐ Non ☐ N/A	

3.3 Infrastructure & Compute

Audit Item	Status	Evidence / Notes
Resource Tracking: GPU/TPU usage and training energy consumption have been estimated or tracked.	☐ Comp ☐ Part ☐ Non ☐ N/A	

Environment Specification: The training and inference environments (Docker containers, dependencies) are fully specified (e.g., requirements.txt, conda.yaml).	☐ Comp ☐ Part ☐ Non ☐ N/A	

3.4 Monitoring Configuration

Audit Item	Status	Evidence / Notes
Drift Detection: Baselines are established for data drift (input distribution) and concept drift (model performance) detection.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Health Metrics: Latency, throughput, and error rate thresholds are defined for production alerting.	☐ Comp ☐ Part ☐ Non ☐ N/A	

4. Operational Governance Checklist

Focus: Maintaining control, accountability, and reliability once the model is live.

4.1 Accountability

Audit Item	Status	Evidence / Notes
Roles Defined: An "Accountable Executive" and "System Owner" are formally designated in the AI Asset Inventory.	☐ Comp ☐ Part ☐ Non ☐ N/A	
RACI Matrix: A RACI matrix exists defining who is Responsible, Accountable, Consulted, and Informed for model issues.	☐ Comp ☐ Part ☐ Non ☐ N/A	

4.2 Incident Management

Audit Item	Status	Evidence / Notes
Incident Response Plan: An AI-specific incident response playbook exists (e.g., for bias detection, performance degradation).	☐ Comp ☐ Part ☐ Non ☐ N/A	
Escalation Path: Clear thresholds exist for when to escalate issues to the AI Ethics Committee or Risk Management.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Kill Switch / Rollback: A technical mechanism exists to immediately take the model offline or roll back to a previous version.	☐ Comp ☐ Part ☐ Non ☐ N/A	

4.3 Lifecycle Management

Audit Item	Status	Evidence / Notes
Retraining Policy: Triggers for retraining (e.g., "accuracy drops below 85%" or "every 3 months") are defined.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Retirement Criteria: Criteria for decommissioning the model are documented (e.g., value drops below cost, new technology available).	☐ Comp ☐ Part ☐ Non ☐ N/A	
Feedback Loop: Mechanisms are in place to collect user feedback/complaints and feed them back into the development cycle.	☐ Comp ☐ Part ☐ Non ☐ N/A	

5. Compliance & Ethics Checklist

Focus: Ensuring the system meets legal obligations and ethical commitments.

5.1 Regulatory Compliance

Audit Item	Status	Evidence / Notes
Regulatory Mapping: Applicable regulations (e.g., EU AI Act, GDPR, NYC LL144, Colorado AI Act) are identified.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Privacy Impact Assessment: A DPIA (Data Protection Impact Assessment) has been completed for systems processing personal data.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Third-Party Audit: If required (e.g., NYC LL144), an independent audit has been scheduled or completed.	☐ Comp ☐ Part ☐ Non ☐ N/A	

5.2 Explainability & Transparency

Audit Item	Status	Evidence / Notes
Explainability Method: Tools (e.g., SHAP, LIME) are implemented to generate local explanations for individual predictions.	☐ Comp ☐ Part ☐ Non ☐ N/A	
User Disclosure: Users are explicitly informed they are interacting with an AI system (if applicable).	☐ Comp ☐ Part ☐ Non ☐ N/A	
Adverse Action: If the system denies services/employment, a mechanism exists to generate specific reason codes for the user.	☐ Comp ☐ Part ☐ Non ☐ N/A	

5.3 Ethical Alignment

Audit Item	Status	Evidence / Notes
Fairness Metrics: Fairness metrics (e.g., demographic parity, equal opportunity) are monitored continuously.	☐ Comp ☐ Part ☐ Non ☐ N/A	
Ethical Review: The system has passed review by the internal AI Ethics Board or equivalent body.	☐ Comp ☐ Part ☐ Non ☐ N/A	

6. Scoring Guide & Next Steps

Scoring Calculation

Count the number of items marked **Compliant** and divide by the total number of **Applicable** items (exclude N/A).

Score: _____ %

Thresholds & Actions

Score Range	Status	Required Action
90 - 100%	Green (Approved)	Proceed with deployment. Schedule next periodic review.
70 - 89%	Yellow (Conditional)	Conditional Approval. Deploy only if no "Critical" items are missing. Establish remediation plan within 30-60 days.
< 70%	Red (Hold)	Do Not Deploy. Significant governance gaps exist. Create remediation roadmap and re-audit before proceeding.

Remediation Timeline

Findings / Gaps Identified	Assigned Owner	Target Resolution Date
1.		
2.		
3.		

Sign-Off

Auditor Signature: _____ **Date:** _____

Executive Sponsor Sign-off: _____ **Date:** _____