



# 2025 GenAI Logging Checklist

🌞 Sandy Dunn



## Who Should Use This Checklist ?

- ★ CISOs, Security Architects, Developers
- ★ Compliance, Privacy, Governance, Legal, Audit Teams
- ★ AI Product Managers AI / ML Engineering & Platform Teams
- ★ DevOps, CloudOps Engineer, Site Reliability Engineers (SREs)







## User Interaction and Model Behavior

What Log Type / Field	Why (Purpose / Description)
<input type="checkbox"/> <b>Prompt Logs</b> (inputs)	Detect misuse, prompt injections, or data leakage attempts
<input type="checkbox"/> <b>Model Output Logs</b> (responses)	Monitor hallucinations, toxic or biased content, sensitive info leakage
<input type="checkbox"/> <b>Session Metadata</b> (User ID, IP, Device, Timestamp, Model Version, Params)	Audit trail and forensic linkage to specific actions
<input type="checkbox"/> <b>Prompt Context</b> Reconstruction	Capture retrieved documents, embeddings, or RAG source IDs
<input type="checkbox"/> <b>Agent Actions &amp; Tool Invocations</b>	Log when agents use external tools or APIs
<input type="checkbox"/> <b>Persona/Prompt Template</b> Versioning	Track changes to system or developer prompts
<input type="checkbox"/> <b>Session Correlation ID</b>	Correlate AI interactions across distributed systems



## Security Monitoring

What Log Type / Field	Why (Purpose / Description)
<input type="checkbox"/> <b>Authentication &amp; Authorization</b> Logs	Prevent unauthorized access or privilege escalation
<input type="checkbox"/> <b>API Gateway</b> Logs	Detects DDoS, scraping, or brute-force activity
<input type="checkbox"/> <b>Third-Party Integration</b> Logs	Monitor plugin or tool access patterns
<input type="checkbox"/> <b>Access Token &amp; Scope Validation</b> Logs	Ensure least-privilege OAuth scope usage
<input type="checkbox"/> <b>Rate Limit</b> Enforcement Logs	Identify abuse or throttling bypasses

<input type="checkbox"/> <b>Sandbox Escape</b> Monitoring	Detects unauthorized system command calls
 <b>Content Risk Monitoring</b>	
What Log Type / Field	Why (Purpose / Description)
<input type="checkbox"/> <b>Toxicity, Bias, and Offensive</b> Output Detection	Prevent reputational or regulatory harm
<input type="checkbox"/> <b>Sensitive Information</b> Detection	Identify leaks of secrets, PII, or confidential data
<input type="checkbox"/> <b>Output Dissemination</b> Tracking	Track propagation of risky or AI-generated outputs
 <b>Data Handling &amp; Storage</b>	
What Log Type / Field	Why (Purpose / Description)
<input type="checkbox"/> <b>Data Uploads / Downloads Logs</b> (for multimodal GenAI: file interactions)	Detects exfiltration or injection attempts via files
<input type="checkbox"/> <b>Encryption Key Access</b> Logs	Track key management and usage
<input type="checkbox"/> <b>Data Retention &amp; Deletion Logs</b> (prompt and output lifespan)	Ensure compliance with data protection regulations like GDPR/CCPA
 <b>Performance &amp; Drift Monitoring</b>	
What Log Type / Field	Why (Purpose / Description)
<input type="checkbox"/> <b>Model Performance</b> Metrics	Track accuracy, latency, and drift over time
<input type="checkbox"/> <b>Explainability Drift</b> Logs	Capture shifts in model reasoning or output attributions
<input type="checkbox"/> <b>Environment Integrity</b> Checks	Detects tampering in deployed model environments
<input type="checkbox"/> <b>Anomalous Output</b> Monitoring (unexpected answer patterns, output structure shifts)	Early detection of model degradation or compromise (eg, poisoned model behavior)
 <b>Incident Detection &amp; Response</b>	
What Log Type / Field	Why (Purpose / Description)
<input type="checkbox"/> <b>Failed Attempts &amp; Errors</b>	Track malformed input and fuzzing attempts

<input type="checkbox"/> <b>Security Event Correlation</b>	Integrate LLM logs into SIEM/SOAR workflows
<input type="checkbox"/> <b>Forensic Replay Logging</b>	Retain anonymized prompt/output pairs for review
<input type="checkbox"/> <b>LLM Runtime Integrity</b>	Hash model files/configs to detect unauthorized modification



## Best Practices

<input type="checkbox"/> <b>Real Time Monitoring</b>	Actively monitor high risk patterns
<input type="checkbox"/> <b>Retention Policies</b>	Define retention, anonymization, and deletion timelines
<input type="checkbox"/> <b>Alerting</b>	Set thresholds for anomalies (eg, more than X failed jailbreaks triggers an alert)
<input type="checkbox"/> <b>Encryption</b>	Log data should be encrypted at rest and in transit
<input type="checkbox"/> <b>Privacy Preserving</b>	Mask/anonymize user data where feasible to comply with privacy laws

## QR For Additional Checklists

