Considering Security, Governance, and Compliance

SECURITY · GOVERNANCE · COMPLIANCE · OWASP LLM TOP 10 · DEFENSE IN DEPTH · SCOPING MATRIX · AWS SERVICES

STUDY & REVIEW

01 MUST KNOW

if you remember five things, remember these

01 **The triad: Security** ensures the confidentiality, integrity, and availability of data, models, and infrastructure. **Governance** establishes the principles, policies, and processes for managing AI systems across their lifecycle. **Compliance** ensures adherence to regulations, industry standards, and internal policies. They are interdependent — address all three to mitigate risk and build trust.

02 **Generative AI raises the stakes:** models are integrated into critical applications and decision-making, process sensitive data, use data that is a **competitive advantage** to its owners, and present **new fronts for adversarial attacks** (OWASP Top 10 for LLMs). AWS calls security **job zero**.

03 **Defense in depth:** implement multiple **redundant** layers of security controls so that if one control fails, other layers still prevent, detect, respond, and recover. Secure everything — workloads, data, accounts, assets — and combine services at every layer.

04 **Know what each service records: AWS Config** = resource configurations and their changes over time · **CloudTrail** = API calls (who did what, when — events) · **CloudWatch** = metrics, monitoring, alarms · **Artifact** = AWS's compliance reports · **Audit Manager** = automated evidence collection against frameworks.

05 **Your data stays yours:** AWS AI infrastructure is **secure by design** — nobody at AWS can access workloads or data running on it, data is encrypted by default, and **no customer data is ever used to train Amazon or third-party foundation models**.

02 KEY TERMS

TERM	DEFINITION
Defense in depth	Layered strategy of redundant security controls applied across every layer of the AI/ML environment
OWASP Top 10 for LLMs	Industry consensus list of the most critical security risks to LLM applications, from the nonprofit OWASP foundation
Prompt injection	Malicious user input crafted to manipulate an LLM's behavior (OWASP #1)
Training data poisoning	Tampering with pre-training, fine-tuning, or embedding data to plant vulnerabilities, backdoors, or bias (#3)
Model theft	Unauthorized access to or exfiltration of a model's parameters or architecture to build a functional copy (#10)
Excessive agency	Granting a model too much autonomy or capability, enabling damaging actions from unexpected outputs (#8)
Overreliance	Trusting authoritative-sounding but erroneous output (hallucination) without oversight or confirmation (#9)
Security Scoping Matrix	AWS framework of five scopes — consumer apps to self-trained models — mapping security duties across the AI lifecycle
Model card	Technical artifact documenting a model's information, strengths, and weaknesses to guide safe application design
Data / concept drift	Over time the model and its training data no longer match reality, degrading predictions

TERM	DEFINITION
Regulated workload	Workload in a highly regulated industry (financial services, healthcare, aerospace) where non-compliance risks health or safety
AWS Audit Manager	Automates evidence collection for compliance assessments; pre-built frameworks include one for generative AI

03 SHARED RESPONSIBILITY FOR GENERATIVE AI

AWS — OF THE CLOUD

security and compliance of the cloud

- Hardware, software, networking, facilities running AWS services
- Purpose-built AI infrastructure — secure by design; nobody at AWS can access it
- Encryption and access controls integrated into AI services; encrypted by default
- Compliance certifications you inherit: ISO 27001, SOC 1/2/3, PCI DSS

CUSTOMER — IN THE CLOUD

security and compliance of your use

- Workloads, data, and applications built on top
- Security groups, access controls, data encryption choices
- Software updates and patches
- Share of work varies with the services you choose — see the scoping matrix

04 OWASP TOP 10 VULNERABILITIES FOR LLMS

know each one-line definition

#	VULNERABILITY	WHAT GOES WRONG
1	Prompt injection	Malicious inputs manipulate the model's behavior
2	Insecure output handling	LLM output passed downstream without validation or sanitization
3	Training data poisoning	Malicious training data teaches harmful behaviors — backdoors, bias
4	Model denial of service	Resource-hungry interactions degrade service quality and drive up cost
5	Supply chain vulnerabilities	Tampered third-party models, training data, software, or platforms
6	Sensitive information disclosure	Outputs leak confidential data or proprietary algorithms
7	Insecure plugin design	Plugins take unvalidated free-text input — up to remote code execution
8	Excessive agency	Too much autonomy → damaging actions from ambiguous outputs
9	Overreliance	Hallucinations trusted without oversight — misinformation, legal risk
10	Model theft	Weights/parameters copied or extracted to clone the model

05 DEFENSE IN DEPTH — SECURE EVERY LAYER

01 **Data** protection → 02 **Identity** & access mgmt → 03 **Application** protection → 04 **Network & edge** protection → 05 **Infrastructure** protection → 06 **Threat detection** & incident response → 07 **Compliance** & governance

06 GENERATIVE AI SECURITY SCOPING MATRIX

responsibility deepens as scope rises

SCOPE	WHAT YOU ARE DOING	EXAMPLE
1	Using public generative AI services as a consumer	PartyRock (an Amazon Bedrock playground); gen-AI apps on the web or phone
2	Using a third-party SaaS service for an enterprise task	Sales management, scheduling, customer-support systems
3	Building an application on a versioned pre-trained model	App on an Amazon Bedrock foundation model

SCOPE	WHAT YOU ARE DOING	EXAMPLE
4	Customizing a model with your data	Fine-tuning with company data
5	Training a model from scratch with your data	Proprietary model, full ML lifecycle

07 AWS SERVICES AT A GLANCE

one-line job of each — exam gold

SERVICE	JOB FOR GENERATIVE AI	REMEMBER
AWS KMS	Centralized encryption-key management; integrates with Bedrock and SageMaker AI	Auto + manual key rotation
Amazon S3	Object storage, encrypted by default; fine-grained IAM access; stores invocation logs, prompts	Server-side encryption
Amazon Macie	ML + pattern matching to discover sensitive data in S3	S3 only · single dashboard
IAM	Least-privilege roles; dev/test/prod separation; per-model Bedrock permissions	Access control
CloudWatch	Real-time monitoring; alerts for prompt injection or disclosure; encrypted logs	Metrics & alarms
CloudTrail	Audit trail of all API calls (CRUD), including Bedrock and S3	Events: who, what, when
AWS Config	Records resource configurations and changes over time; rules evaluate compliance	Setup: what & how configured
CAF for AI/ML/gen AI	Governance perspective: cost, data curation, risk, responsible AI	Six perspectives total
AWS WA Tool	Measures architecture against Well-Architected best practices	Includes an ML lens
Trusted Advisor	Inspects your environment: save money, improve availability/ performance, close security gaps	Recommendations
Amazon Inspector	Automated security assessments; findings prioritized by severity	Pipeline integration
AWS Artifact	On-demand downloads of AWS compliance reports and agreements	ISO, SOC, PCI DSS · ISV reports
AWS Audit Manager	Automates evidence collection; pre-built compliance frameworks	Gen-AI framework: 8 principles

08 EXAM TRAPS

where the knowledge check gets you

- × "Compliance is a box you check once." — Compliance is **continuous**: risk assessment, auditing, monitoring, and revising policies as results come in.
- × "AWS Config tracks API calls." — **CloudTrail** records API calls (events — who did what, when). **Config** records how resources are configured and how that changes.
- × "AWS uses customer data to improve its foundation models." — **Never.** Regardless of the service, no customer data is used to train Amazon or third-party FMs.
- × "Under shared responsibility, AWS handles compliance for what you build." — AWS is responsible for compliance **of** the cloud; you are responsible for compliance of your **use** of AWS services.
- × "Amazon Inspector tracks resource configuration changes." — Inspector runs **security assessments** (vulnerabilities, exposures, deviations from best practices). Configuration history is **AWS Config**.
- × "AWS Artifact audits your environment." — Artifact only serves **AWS's own** compliance reports and agreements for download; it never scans your resources.
- × "Every generative AI app needs the same security." — Depth and breadth scale with use case and data: a movie finder on PartyRock vs. a healthcare app on a proprietary model are worlds apart.
- × "Macie protects sensitive data across all AWS storage." — Macie discovers and helps protect sensitive data **in Amazon S3** specifically.

09 SELF-QUIZ

answers are upside-down below

Q1 Which function of the triad establishes the principles, policies, and processes for managing AI systems throughout their lifecycle?

Q2 Security ensures which three properties of data, models, and infrastructure?

Q3 Which OWASP LLM vulnerability manipulates a model's behavior through malicious user input?

Q4 Which OWASP LLM vulnerability involves extracting or copying a model's parameters to create a functional equivalent?

Q5 What layered security strategy ensures that if one control fails, others still prevent, detect, respond, and recover?

Q6 Which AWS service uses machine learning and pattern matching to automatically discover sensitive data in S3 buckets?

Q7 In the Generative AI Security Scoping Matrix, what does Scope 5 cover?

Q8 Which AWS service maintains the audit trail of API calls — who, what, when — across services such as Amazon Bedrock and S3?

Q9 Which AWS service automates evidence collection and provides a pre-built best-practices framework for generative AI?

Q10 Sample exam: a company needs to track changes to its resources for compliance — Amazon Inspector, AWS Config, AWS Artifact, or Amazon CloudWatch?

ANSWER KEY (rotate the page)

Q1 governance **Q2** confidentiality, integrity, availability **Q3** prompt injection (#1) **Q4** model theft (#10) **Q5** defense in depth **Q6** Amazon Macie
Q7 training a model from scratch with your own data **Q8** AWS CloudTrail **Q9** AWS Audit Manager **Q10** AWS Config