Session 4 Handout: Responsible AI, Security, and Governance (Domains 4 and 5, 14% each)

FAST-12 AI Practitioner Certification Prep · 7formations · Instructor: Wadood Chaudhary · All examples fictional · Exam guide v1.2

Why this session matters

Together these domains are 28% of the exam. Questions ask which dimension a risk falls under and which service or control addresses it.

Responsible-AI dimensions (AWS)

Fairness, explainability, privacy and security, safety, controllability, veracity and robustness, governance, transparency.

Risk (fictional)	Dimension	Control
Clinic chatbot gives medication advice	Safety, controllability	Guardrail denied topic; human handoff
Loan model approves fewer applicants from one ZIP code	Fairness	Bias metrics on subgroups; review features
No one can say why a claim was flagged	Explainability	Feature attributions; simpler model where needed
Users don't know they're talking to AI	Transparency	Disclosure; model cards

Amazon Bedrock Guardrails: content filters, denied topics, word filters, sensitive-information filters (mask or block PII), contextual grounding checks, Automated Reasoning checks. **Human review** for high-impact decisions. **SageMaker Model Cards** document intended use and limits. Note: SageMaker Clarify closed to new customers in 2026; older guides still mention it.

Security: shared responsibility

AWS secures **the cloud** (infrastructure, managed services). You secure **in the cloud**: data, IAM, configuration, encryption choices, application, guardrails.

Need	Service
Least-privilege access	IAM
Encryption keys	AWS KMS
Audit API calls	AWS CloudTrail
Monitor metrics and logs	Amazon CloudWatch
Track configuration compliance	AWS Config
Find sensitive data in S3	Amazon Macie

Need	Service
Compliance reports, agreements (BAA)	AWS Artifact
Vulnerability scanning	Amazon Inspector
Store secrets	AWS Secrets Manager
Private connectivity	Amazon VPC (endpoints)

Governance

Data lineage and cataloging (AWS Glue, Lake Formation), retention policies, model and prompt versioning, audit cadence, AWS Well-Architected Tool reviews, frameworks such as the NIST AI RMF and ISO/IEC 42001. Massachusetts personal information (G.L. c. 93H, 201 CMR 17.00) and PHI (HIPAA) add legal obligations beyond the exam.

Checklist

- ☐ I can map a risk to a responsible-AI dimension.
- ☐ I can configure the parts of a Bedrock guardrail.
- ☐ I can split a scenario's controls into AWS vs customer.
- ☐ I can match each security need to its service.