Day 1

Foundations of AI Governance:

What Is AI Governance?

AI Governance means setting the rules, processes, and responsibilities that ensure AI systems are safe, ethical, reliable, and compliant with laws. It is not about building AI. It is about managing how AI is designed, tested, used, monitored, and controlled.

Objectives of AI Governance:

1. Reduce Risks
2. Ensure Ethical & Fair AI
3. Improve Transparency
4. Ensure compliance with laws
5. Maintain Accountability
6. Improve Trust and reliability

Difference Between AI Governance and AI Security:

AI Governance	**AI Security**
Rules and processes for *safe + ethical* use of AI	Protecting AI from *attacks and misuse*
Focus on ethics, fairness, transparency, compliance	Focus on threats, vulnerabilities, hacking risks
Like “policies & management”	Like “technical protection”
People + processes	Tools + defenses

Governance as a Component of Responsible AI:

Responsible AI (RAI) means building AI that is fair, ethical, safe, inclusive, and trustworthy. AI Governance is a core part of Responsible AI because it creates the system that enforces:

- Fairness rules
- Security requirements
- Bias checks
- Transparency
- Human oversight

- Model monitoring
- Ethical use

Scope of AI Governance: AI governance controls everything from input → training → deployment → monitoring.

Importance of Governance in AI Security:

Governance makes AI security consistent, predictable, and enforceable across an organization.

How Governance Reduces AI Risks?

Governance reduces AI risks by placing structured controls around every stage of an AI system.

Area	What Governance Requires	What It Checks	Why It Matters	Without Governance
1. Risk Assessments	Mandatory risk reviews for every model	Data use, bias, attack vectors, failure impact	Identifies risks early and ensures models are safe and ethical	Teams skip reviews → risks rise significantly
2. Approval Processes	Testing, verification, documentation, and proper approvals	Validation of model performance and completeness of documentation	Prevents untested or weak models from being deployed	Poor-quality or unsafe models enter production
3. Standardization	Common rules for security, documentation, testing, and monitoring	Consistency of practices across all teams	Reduces human error and improves reliability	Each team does things differently → mistakes increase
4. Ongoing Monitoring	Continuous evaluation of models after deployment	Drift, accuracy drops, unusual behavior, suspicious API usage, data issues	Detects problems early, before they cause harm	Issues go unnoticed until they cause major failures
5. Clear Responsibilities	Defined ownership for creation, security, approval, monitoring, and incident response	Accountability structures	Ensures someone is always responsible for model health and safety	Confusion over roles → slow response and avoidable mistakes

Role of Governance in Preventing Attacks:

Governance plays a non-technical but powerful role in preventing attacks on AI systems.

1. Defense against Data Poisoning
2. Defense against Prompt injection
3. Prevent model misuse
4. Model Security Requirements

Governance for Regulatory Compliance

- Many regulations govern AI (e.g., EU AI Act, DPDP Act, NIST AI RMF, ISO 42001).
- Governance helps ensure the organization stays compliant with all legal requirements.

How governance supports compliance:

- Keeps audit-ready documentation
- Maintains risk logs
- Enforces data protection measures
- Classifies models by risk level
- Assigns clear accountability for each model
- Requires human oversight for high-risk systems
- Implements processes for explainability, consent, and user appeals

Without governance:

- Companies may unintentionally violate laws and face penalties.

Governance for Ethical and Safe Deployment:

Governance ensures AI is deployed carefully, not blindly.

Governance Impact on Incident Response and Monitoring

Governance brings discipline to incident response. It strengthens incident response by providing playbooks, logs, monitoring rules, and quick recovery plans.

--The End--