From Principles to Power:
The 2015–2025 AI Governance Turn
Laws, Standards, and What to Do Next

Guest Lecture by Adnan Masood, PhD.

Graduate Seminar on AI Governance

October 8, 2025

How to use this deck (interactive format)

Structure per instrument: *Title → Definition/Scope → Intuition/Example → Obligations/Enforcement/Timeline*.

Class prompts: discussion questions are embedded on many slides for think-pair-share.

Focus: Only the regulations, statutes, and standards covered in the assigned reading.

What we skip: generic definitions ("audit", "governance") unless essential to a rule.

Goal: equip you to map use-cases to obligations across jurisdictions and standards.

Agenda

Roadmap

First intergovernmental AI framework adopted by 40+ countries; later echoed by G20.

OECD AI Principles: Definition & Scope

Instrument: Non-binding Recommendation setting 5 value-based principles (*inclusive growth; human-centered values; transparency; robustness; accountability*) and 5 policy recommendations.

Scope: Applies to AI systems broadly (definition influenced later laws). Adopted by OECD members and adherents.

Purpose: Common baseline to guide national policy, standards, and cross-border cooperation.

OECD AI Principles: Intuition & Example

Intuition: Create a shared lingua franca for "trustworthy AI" before binding laws emerge.

Example: A national AI strategy cites OECD principles to justify transparency obligations for recommender systems.

Prompt: Where do these principles map to *binding* duties in later regimes?

Obligations: None directly; influential in procurement criteria and regulator guidance.

Enforcement: Soft law (peer pressure; policy reviews via OECD AI Policy Observatory).

Timeline: Adopted 2019; continuing implementation via national roadmaps and OECD tooling.

First universal (193 states) normative instrument on AI ethics.

UNESCO Recommendation: Definition & Scope

Instrument: Non-binding recommendation addressing data governance, human rights, inclusion, environmental impact.

Scope: Governments commit to implement via domestic policy; calls for safeguards against rights-violating AI uses.

Interfaces: Informs national AI strategies, public-sector policies, and education.

UNESCO Recommendation: Intuition & Example

Intuition: A human-rights-centered anchor to shape emerging national rules.

Example: A country adopts a moratorium on some biometric surveillance citing UNESCO guidance.

Obligations: Programmatic—impact assessment tools, oversight mechanisms, bans for rights-incompatible uses (via domestic law).

Enforcement: Moral suasion; periodic monitoring by UNESCO.

Timeline: Adopted Nov 2021; national implementations vary 2022–2025.

Multilateral commitments on generative/frontier AI safety and cooperation.

Instrument: Joint statements and principles on risk-based governance, evaluation, information sharing.

Scope: Voluntary; targeted at generative and "frontier" models and platform risks.

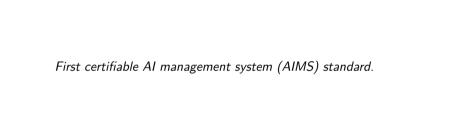# G7 & Bletchley: Intuition & Example

Intuition: Rapid coordination for fast-moving foundation models; prepare evaluation/red-teaming norms.

Example: Safety institutes (UK-led) test models; shared workstreams on evaluations.

Obligations: None legally binding; influence procurement and national standards.

Timeline: 2023 summits; follow-ons in 2024/2025 iterate on evaluation guidance.

First certifiable AI management system (AIMS) standard.

Instrument: International standard specifying requirements for an AI management system (policy, roles, risk processes).

Scope: Organizational-level; can be certifiable by accredited bodies.

Fit: Complements legal requirements (e.g., EU AI Act quality mgmt system).

Intuition: "ISO 9001 for AI": institutionalize governance so each AI system follows a controlled lifecycle.

Example: A bank certifies to 42001 to show due diligence to regulators and customers.

ISO/IEC 42001: Obligations, Enforcement & Timeline

Obligations: Establish AIMS, risk criteria, documentation, competence, monitoring, continual improvement.

Enforcement: Voluntary; becomes quasi-mandatory via contracts or regulatory recognition.

Timeline: Published 2023; adoption accelerates 2024–2027 alongside EU AI Act roll-out.

Process standard aligned to ISO 31000 for AI-specific risks.

ISO/IEC 23894: Definition & Scope

Instrument: International guidance on identifying, analyzing, evaluating, and treating AI risks.

Scope: System-level risk processes across lifecycle (data, model, deployment, post-market).

ISO/IEC 23894: Intuition & Example

Intuition: A common, auditable approach to bias, robustness, safety, and explainability risk.

Example: Impact assessment template uses 23894 controls as checklist.

ISO/IEC 23894: Obligations, Enforcement & Timeline

Obligations: Voluntary; often paired with NIST AI RMF and sector rules.

Timeline: 2023 publication; referenced by regulators and harmonized standards under EU AI Act.

Transparency of autonomous systems; algorithmic bias considerations.

IEEE 7001/7003: Definition & Scope

Instrument: Technical standards specifying transparency artifacts (7001) and bias controls (7003).

Scope: Design-time and evaluation-time requirements; complements law.

Intuition: Make "explainability" and "bias" tangible and testable.

Example: A vendor supplies a 7001 transparency statement and 7003 bias test report with their model card.

IEEE 7001/7003: Obligations, Enforcement & Timeline

Obligations: Voluntary; may be baked into procurement and audits (e.g., NYC hiring tools).

Timeline: Published 2021/2022; adoption grows via assurance ecosystems.

Roadmap

Horizontal, risk-tiered law with GPAI/foundation-model duties; extraterritorial reach.

EU AI Act: Definition & Scope

Instrument: EU Regulation (directly applicable) with four tiers: prohibited, high-risk, limited-risk (transparency), minimal.

Scope: Providers, deployers, importers placing AI on EU market or whose outputs affect people in the EU.

High-risk set: Safety components of regulated products and Annex III domains (e.g., employment, credit, essential services, law enforcement).

EU AI Act: Intuition & Example

Intuition: Product-safety style, ex-ante compliance to protect fundamental rights and safety.

Example: An HR-screening system in Annex III must undergo conformity assessment and be registered before use.

Core duties (high-risk): risk mgmt system; data governance; technical documentation; human oversight; accuracy/robustness/cybersecurity; CE marking; post-market monitoring.

GPAI/foundation models: documentation on capabilities and training-data summaries; safeguards for illegal content; additional duties for very capable models.

Enforcement: National market surveillance + EU AI Office; fines up to 6% global turnover.

Timeline (phased): bans early; GPAI obligations ∼12 months post-entry; high-risk ∼24 months; legacy grace windows.

Automated decision-making limits and rights; DPIAs for high-risk processing.

GDPR ADM: Definition & Scope

Instrument: EU data protection law provision on decisions based solely on automated processing with significant effects.

Scope: Requires legal basis and offers right to obtain human intervention, to express viewpoint, to contest decision.

Interface: Continues to apply alongside the AI Act.

Intuition: Preserve agency where algorithms decide consequential outcomes.

Example: Credit denial via automated scoring must provide meaningful information and human review path.

GDPR ADM: Obligations, Enforcement & Timeline

Obligations: Inform data subjects; ensure fairness; conduct DPIA for high-risk profiles; honor access/explanation rights (where applicable).

Enforcement: Data Protection Authorities; fines under GDPR regime.

Timeline: In force since 2018; active enforcement through DPAs.

Algorithmic accountability for very large online platforms (VLOPs) and transparency to users.

DSA: Definition & Scope

Instrument: EU Regulation imposing risk assessments, independent audits, recommender transparency and user choice (including non-profiled feed).

Scope: Platforms meeting VLOP thresholds; general duties across intermediaries.

DSA: Intuition & Example

Intuition: Mitigate systemic risks amplified by recommender algorithms (disinformation, harms to minors).

Example: A VLOP must document risk mitigations for recommender systems and offer a non-profiling option.

DSA: Obligations, Enforcement & Timeline

Obligations: Annual systemic risk assessments; transparency reporting; access to data for vetted researchers; labeling for manipulated media.

Enforcement: European Commission for VLOPs; national coordinators otherwise.

Timeline: Effective 2023 for VLOPs; full application 2024.

Ease victim redress for AI-caused harm; clarify software as product.

EU Liability: Definition & Scope

Instrument: AI Liability Directive (proposal) and revised Product Liability Directive.

Scope: Presumptions of causality in certain non-compliance; disclosure obligations; extends strict liability to software and updates.

EU Liability: Intuition & Example

Intuition: Close proof gaps where opaque AI complicates causation.

Example: Court can order disclosure of technical docs from provider to injured party.

Status: Advancing in legislative process; expected to complement AI Act enforcement window.

Impact: Increases litigation exposure; elevates documentation and logging importance.

Roadmap

Record-filing, transparency, and user controls for recommender systems.

Instrument: CAC-led binding rules for algorithmic recommendation services.

Scope: Disclosure of basic principles, purposes, and mechanisms; provide opt-out and non-targeting options; prohibit illegal data use and harmful amplification.

Registration: Algorithm record-filing with CAC for certain services.

Intuition: Align platforms with state policy goals and reduce harmful social effects of algorithms.

Example: Platform offers "no personalization" toggle and publishes algorithm filing ID.

Obligations: Filing; transparency; content and data-use constraints; complaint handling.

Enforcement: CAC inspections, orders, fines; public registry.

Timeline: Effective March 1, 2022.

Labeling and controls for AI-generated (deepfake) content.

Instrument: Binding rules requiring clear labeling of AI-generated/altered media; user identity verification; log retention.

Scope: Providers of deep synthesis tools and services to the public.

Intuition: Prevent deception and misuse of AI media; ensure traceability.

Example: Video app watermarks AI-altered clips and keeps processing logs.

China Deep Synthesis: Obligations, Enforcement & Timeline

Obligations: Prominent labeling; security assessments; misuse handling.

Timeline: Effective Jan 10, 2023; enforcement via CAC & other agencies.

Security reviews, content controls, algorithm filing, and output labeling.

Instrument: Binding interim measures covering public generative services.

Scope: Accuracy and legality of content; safeguards to reflect "core socialist values"; registration and assessments.

Intuition: License-and-control regime for public genAI amid rapid deployment.

Example: Chatbot provider performs model updates after generating illegal content and reports incident to regulator.

Obligations: Algorithm filing; security assessment; watermarking/labeling; content remediation; user real-name verification.

Enforcement: CAC and co-regulators; corrective orders and fines.

Timeline: Effective Aug 15, 2023.

Roadmap

Whole-of-government directives on safety standards, procurement, and frontier model oversight.

Instrument: Executive policy instructing agencies (e.g., NIST, DHS, HHS, DoL) to set AI safety, security, and civil-rights guardrails.

Scope: Federal use and procurement; notification for very large training runs; watermarking guidance; worker protections.

Intuition: Use executive power to coordinate in absence of omnibus federal statute.

Example: Federal contractors asked to align to NIST AI RMF and safety test guidance.

Obligations: Agency actions; standards and reporting that cascade to vendors.

Enforcement: Administrative (procurement leverage; OMB oversight).

Timeline: Issued Oct 2023; rolling agency deliverables through 2024–2025.

US voluntary framework: Govern, Map, Measure, Manage.

Instrument: Consensus guidance with Playbook; profiles for sectors and genAI emerging.

Scope: Any organization developing or using AI; complements ISO 23894/42001.

Intuition: Practical scaffolding for trustworthy AI without prescriptive law.

Example: A deployer builds an AIA aligned to RMF functions with metrics for bias and robustness.

Obligations: Voluntary; often required in federal procurement and state laws' safe harbors.

Timeline: v1.0 Jan 2023; continuing updates and profiles.

Bias audits and notices for automated employment decision tools (AEDTs).

Instrument: City law mandating annual independent bias audits and candidate notices for AEDTs used in NYC.

Scope: Hiring and promotion tools that substantially assist automated decisions.

Intuition: Transparency and disparate-impact checks in employment where harms are acute.

Example: Vendor publishes public audit summary with selection rate ratios across demographics.

NYC LL 144: Obligations, Enforcement & Timeline

Obligations: Pre-use audit; candidate notice and opt-out; data retention for auditing.

Enforcement: City Department of Consumer and Worker Protection; penalties per violation.

Notice, explanation, consent for AI analysis of interview videos; reporting.

Instrument: State statute governing use of AI to evaluate video interviews.

Scope: Employers using AI analysis in hiring; obligations around consent and data handling.

Intuition: Early, narrow use-case law to protect applicants.

Example: Employer must explain how AI evaluates candidates and obtain express consent.

Obligations: Notice, explanation, consent; delete videos upon request; demographic reporting pilot.

Enforcement: State AG; private remedies depend on context.

Bias prohibitions in insurance; reasonable care, impact assessments, and transparency for high-risk AI.

Instrument: Sectoral rule for insurance unfair discrimination; broad AI statute imposing duties on developers and deployers of high-risk AI.

Scope: High-risk = significant risk of discrimination or significant effects on consumers (e.g., credit, employment, housing, insurance).

Intuition: Blend of consumer protection and civil-rights logic applied to AI outcomes.

Example: Insurer maintains governance program and bias testing under SB 21-169; software provider performs impact assessment and publishes summary under SB 24-205.

Obligations: Reasonable care to avoid algorithmic discrimination; impact assessments; incident reporting; transparency to consumers interacting with AI.

Enforcement: State Attorney General; safe-harbor concepts via documented risk programs.

Timeline: 2021 (insurance regs phased in 2023); 2024 AI Act with staged dates.

Roadmap

Principles-first, regulator-led approach (no omnibus Act yet).

Instrument: Policy setting five cross-sector principles (safety, transparency, fairness, accountability, contestability).

Scope: Implemented via existing regulators (ICO, CMA, FCA, MHRA, etc.).

Intuition: Agile, innovation-friendly governance via guidance and sector supervision.

Example: ICO issues explainability guidance for ADM under UK GDPR; CMA principles for foundation models competition.

Obligations: Non-binding principles; regulators incorporate into expectations and codes.

Related law: Online Safety Act 2023 imposes algorithmic risk duties for large platforms.

Roadmap

High-impact AI obligations; AI & Data Commissioner oversight.

Canada AIDA: Definition & Scope

Instrument: Proposed federal law as part of C-27; focuses on "high-impact" AI designated by regulation.

Scope: Risk assessments, mitigation, monitoring, incident reporting; prohibits harmful intent uses.

Intuition: EU-like focus on high-impact uses with flexible designation via regulations.

Example: Provider documents risks and mitigation for an AI used in credit decisions; keeps records for audits.

Enforcement: AI & Data Commissioner; administrative and criminal penalties for egregious breaches.

Status: Passed House (2023); under Senate review through 2025.

Mandatory Algorithmic Impact Assessment (AIA) for federal services.

Canada ADM Directive: Definition & Scope

Instrument: Treasury Board directive; binding on federal departments.

Scope: Impact levels I–IV drive controls (peer review, source code openness, human oversight); public inventory of systems.

Intuition: Governance-by-procurement and transparency for public-sector ADM.

Example: Immigration triage tool undergoes AIA and publishes overview.

Canada ADM Directive: Obligations, Enforcement & Timeline

Obligations: AIA; documentation; human-in-the-loop at higher levels; notices to affected individuals.

Enforcement: Treasury Board oversight and audits.

Promotional law with transparency/risk duties for high-impact AI; effective 2026.

Instrument: National framework balancing innovation and trustworthy AI.

Scope: "High-impact" categories (e.g., health, finance); labeling for AI-generated content in contexts; coordination across ministries.

Korea AI Framework: Intuition & Example

Intuition: Moderate guardrails with relatively light penalties; administrative guidance-forward.

Example: Financial AI adheres to transparency guidance; deployer labels AI content per sector rules.

Korea AI Framework: Obligations, Enforcement & Timeline

Obligations: Risk management and transparency for designated AI; fines at administrative scale.

Timeline: Promulgated Jan 2025; enters into force Jan 2026.

Principle-based, non-penal statute to steer policy and coordination.

Instrument: Framework law promoting AI R&D and responsible use; establishes strategic council.

Scope: No direct private-sector obligations; leverages existing laws and guidelines.

Japan AI Act: Intuition & Example

Intuition: Keep innovation pace high while using soft-law to manage risks.

Example: Industry follows METI/MIC governance guidelines aligned with OECD principles.

Obligations: Voluntary adherence; sector regulators issue guidance as needed.

Timeline: Effective Sept 2025.

Detailed voluntary guidance and a testing toolkit for AI governance.

Instrument: Model governance framework (v2/v3) and AI Verify testing suite.

Scope: Practical controls (ops mgmt, stakeholder communication, transparency); pilot assurance reports.

Intuition: Market-led assurance that can scale into certification; procurement lever.

Example: Vendor provides AI Verify report with claims and test evidence.

Obligations: Voluntary; increasingly referenced in public-sector procurements.

Roadmap

Toward risk-based obligations and an oversight body; building on LGPD ADM rights.

Brazil AI Bill: Definition & Scope

Instrument: Pending legislation; Senate draft strengthens duties and rights incl. explanations.

Scope: Critical/high-risk use-cases flagged; complements LGPD Art. 20 automated decision rights.

Brazil AI Bill: Intuition & Example

Intuition: EU-inspired structure adapted to Brazil's constitutional and consumer law context.

Example: Credit-scoring providers required to provide logic information and redress channels.

Status: Under debate; expect phased implementation after passage.

Roadmap

GDPR-style right to human review for solely automated, consequential decisions.

Kenya ADM Rights: Definition & Scope

Instrument: Data protection statute with ADM clauses.

Scope: Individuals can request human intervention and challenge automated decisions.

Intuition: Rights-forward approach to algorithmic decisions in emerging markets.

Example: Fintech lending app provides manual review path for adverse decisions.

Enforcement: Office of the Data Protection Commissioner; complaints and audits.

Transparency and challenge rights for impactful automated decisions.

Instrument: National data protection law modernizing ADM controls.

Scope: Notice and appeal routes for solely automated, significant decisions.

Nigeria ADM Rights: Intuition & Example

Intuition: Align with global privacy norms; rein-in opaque scoring tools.

Example: Mobile credit scoring subject to user contestation and oversight.

Enforcement: Nigeria Data Protection Commission; guidance emerging.

Roadmap

First binding human-rights-centered AI treaty (pending adoption).

CoE AI Convention: Definition & Scope

Instrument: Multilateral treaty focusing on human rights, democracy, rule of law safeguards for AI.

Scope: Public sector and possibly private when performing public functions; risk assessments; oversight.

Intuition: Create a transnational floor for rights-preserving AI governance across Europe (EU and non-EU).

Status: Text finalized in committee; open for signature expected 2024/25; domestic ratifications to follow.

Roadmap

Comparative models at a glance

	EU (AI Act)	US (EO/NIST & States)	China (CAC rules)
Style	Risk-tiered, prescriptive	Sectoral, enforcement-led, voluntary standards	Category-specific, licensing/content control
Scope	Providers & deployers, extraterritorial	Agencies & sectors; state patchwork	Public-facing platforms/services
Duties	Ex-ante conformity, docs, oversight	Impact audits (NYC), reasonable care (CO), notices	Filing, labeling, security reviews, value alignment
Teeth	Up to 6% turnover	AG/FTC/CFPB enforcement; civil rights	Administrative orders; fines; registration

EU AI Act ↔ *ISO/IEC 42001* (quality management) & *ISO/IEC 23894* (risk), plus IEEE 7001/7003 for artifacts.

US NIST AI RMF ↔ maps cleanly to ISO 23894 functions; often used as practical playbook.

NYC LL 144 ↔ IEEE 7003 bias considerations; independent audit reporting patterns.

Roadmap

You deploy an AI screening tool in NYC and across the EU.

Discuss: What documents and tests must exist *before* go-live? Who signs off? What public disclosures are required?

Hint: NYC LL 144 (audit+notice) + EU AI Act (Annex III HR high-risk) + GDPR ADM rights.

You add an image generator and chatbot available globally.

Discuss: Labeling/watermarking, security reviews, content filters, and timeline differences for EU (transparency), US (voluntary/NIST), China (filing and content rules).

Discuss: SB 21-169 unfair discrimination testing; EU Act high-risk under essential services; documentation and monitoring regimes.

Roadmap

Minimal viable AI governance program (reg-driven)

Inventory & classify: map systems to jurisdictional obligations (EU high-risk, NYC AEDT, CO high-risk).

Impact/risk assessments: adopt NIST AI RMF/ISO 23894 aligned templates; store technical files.

Human oversight: define roles and escalation paths per use-case.

Bias & robustness testing: pre-deployment and periodic; vendor attestations; logs for audits.

Transparency: user notices; model cards; deepfake labeling where required.

Post-market monitoring: incident reporting triggers; regression guardrails for updates.

Roadmap

Key dates (selected, 2024–2027)

EU AI Act: bans and governance early; GPAI ∼2025; high-risk ∼2026; full ramp by 2027.

China: recommender (2022) → deep synthesis (2023) → genAI (2023) in force.

US: EO deliverables rolling; NYC LL 144 in force 2023; CO AI 2024 staged.

Canada: AIDA progress through 2025.

Korea: Framework Act in force 2026.

Discussion prompts to take home

Where do standards create *de facto* safe harbors across regimes?

What is the strongest extraterritorial hook you must plan for in your product?

How would you evidence "reasonable care" if audited tomorrow?

Summary

Convergence on core duties: risk assessment, documentation, transparency, human oversight, monitoring.

Divergence in structure: EU prescriptive tiers; US sectoral + state patchwork; China licensing/control; UK/Japan principle-led.

Practical path: build to the strictest applicable regime; evidence through standards-aligned artifacts.

Q & A