INDRC

AI SUMMER SCHOOL 2026

Responsible and Trustworthy AI: Challenges, Compliance, and Best Practices

Operationalizing Core Legal Standards Under EU Regulation (EU) 2024/1689

PART 1: RESPONSIBLE AI
Ethical Pillars & Legal Codification

PART 2: CHALLENGES
Risk Taxonomy & Statutory Red Lines

PART 3: COMPLIANCE
Ex-Ante Requirements & Fine Tiers

PART 4: BEST PRACTICES
Regulatory Sandboxes & Roadmap

Introduction: Harmonizing Innovation & Fundamental Rights

1. Legislative Purpose

- Harmonize internal EU market rules for AI systems.
- Protect health, safety, and fundamental rights (EU Charter).
- Foster investment & innovation in human-centric AI.
- Establish extra-territorial reach for global providers.

2. Scope & Key Exclusions

- Applies to: EU providers, deployers, and 3rd-country outputs used in EU.
- Excludes: Military, defense, and national security systems.
- Excludes: Pure scientific research & development activities.
- Excludes: Personal, non-professional usage by individuals.

3. Entry Into Force Timeline

- August 2024: Act enters into force across Union.
- February 2025: Article 5 Prohibited Practices take effect.
- August 2025: General-Purpose AI (GPAI) model rules apply.
- August 2026: Full application for Annex III High-Risk systems.

Part 1: Question — What Makes AI Systematically Trustworthy?

THE CORE QUESTION FOR AI DEVELOPERS:

“Can autonomous algorithmic decision-making earn human trust without statutory legal safeguards?”

Dilemma A: The Capability vs Transparency Paradox

- Deep neural networks achieve state-of-the-art predictive accuracy.
- However, their complex parameters lack human-interpretable explainability.
- Opacities lead to undetected algorithmic bias and loss of public confidence.

Dilemma B: The Limits of Voluntary Corporate Ethics

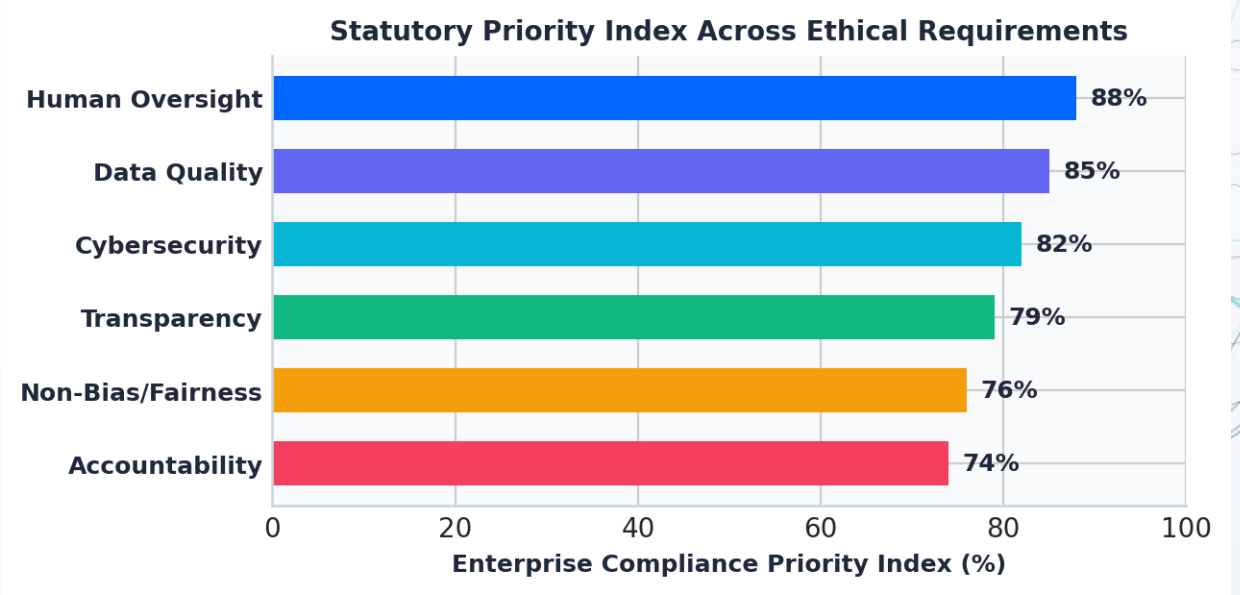
- Self-regulatory corporate guidelines fail without statutory enforcement.
- Market incentives often prioritize rapid deployment over safety audits.
- Harmonized legal standards are necessary to level the playing field.

Part 1: Data — Codifying Ethical Principles into Statutory Law

PART 1: DATA

7 HLEG Ethical Pillars Codified (Articles 8–15)

1. Human Agency & Oversight (Art. 14): Mandatory stop buttons.
2. Technical Robustness & Safety (Art. 15): Resilience against attacks.
3. Data Governance & Bias (Art. 10): Representative datasets.
4. System Transparency (Art. 13): Clear instructions for use.
5. Diversity & Fairness (Art. 10): Proactive bias mitigation.
6. Environmental Well-being: Resource & energy tracking.
7. Verifiable Accountability (Art. 11/12): Audit logs & dossiers.



Part 1: Example — Real-World Algorithmic Trust Failures

PART 1: EXAMPLE

Case 1: Amazon ML Recruitment Tool Gender Bias

- Incident: Machine learning resume screening tool trained on 10 years of historical applicant data.
- Failure: Algorithmic model penalized resumes containing the word 'women's' (e.g., 'women's chess club captain').
- Outcome: Project shut down; highlighted risks of historical training bias.
- Statutory Fix: Article 10 mandates representative, error-free dataset curation and proactive bias testing.

Case 2: Dutch SyRI Welfare Profiling Litigation (2020)

- Incident: System Risk Indication (SyRI) algorithm calculated welfare fraud risk scores across low-income neighborhoods.
- Failure: Black-box risk scoring violated privacy and disproportionately targeted socio-economically vulnerable citizens.
- Outcome: Hague District Court invalidated SyRI for violating Article 8 ECHR.
- Statutory Fix: Article 5 social scoring bans and Article 27 Fundamental Rights Impact Assessments (FRIA).

Part 1: Answer — The Statutory Engine for Human-Centric AI (Operational)

PART 1: RESOLUTION

1. Human Oversight

- Technical stop buttons & dual-person verification controls (Art. 14).

2. Data Governance

- Bias auditing, error-free curation, representative datasets (Art. 10).

3. Transparency

- Clear deployer manuals, limitations, & accuracy metrics (Art. 13).

4. Accountability

- Technical documentation, 6-mo logs, CE marking (Art. 11/12).

Part 2: Question — Where Must Law Draw Statutory Red Lines?

THE STATUTORY RISK QUESTION:

“When does an AI application cross the line from beneficial innovation into unacceptable societal harm?”

Risk A: Biometric Surveillance & Civil Liberties

- Untargeted mass facial recognition erodes public anonymity.
- Predictive policing profiling creates self-fulfilling bias loops.
- Real-time biometric tracking risks chilling freedom of assembly.

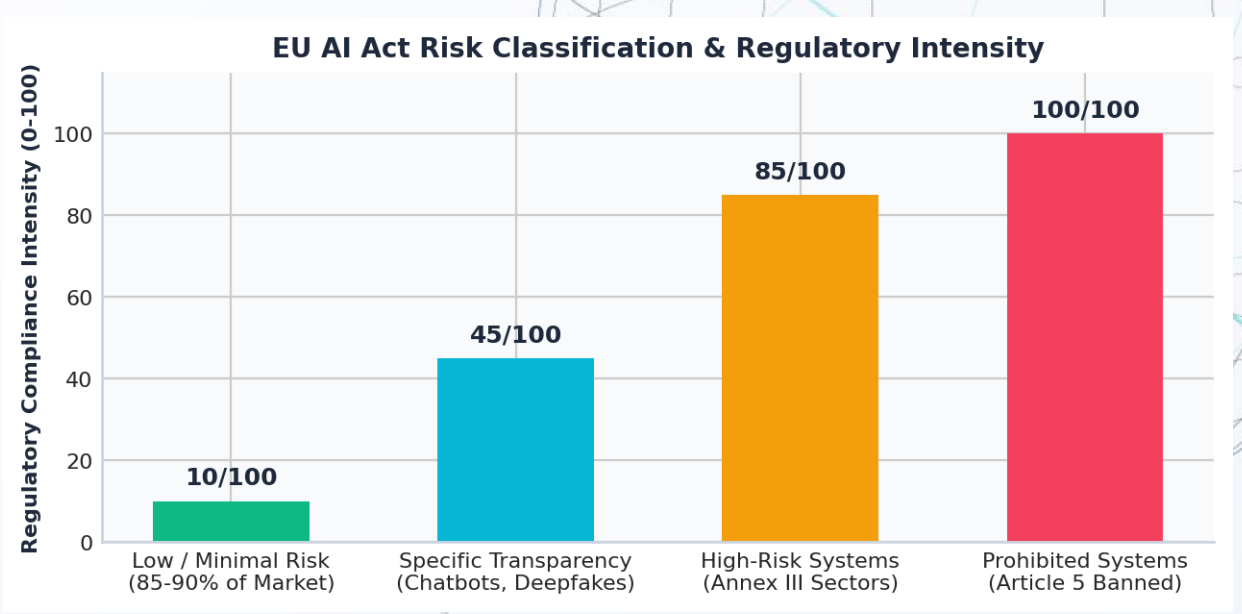
Risk B: Subliminal Manipulation & Exploitation

- AI systems exploiting cognitive vulnerabilities of minors or vulnerable groups.
- Social scoring classifying citizens based on unrelated social behavior.
- Emotion recognition inferring inner states in employment and schools.

Part 2: Data — Article 5 Prohibited AI Practices

8 Statutory Banned Practices (Article 5)

1. Subliminal / Deceptive Behavioral Distortion (Art. 5(1)(a))
2. Exploitation of Age, Disability, or Socio-Economic Status (Art. 5(1)(b))
3. Social Scoring Leading to Unfavorable Treatment (Art. 5(1)(c))
4. Predictive Policing Profiling (Art. 5(1)(d))
5. Untargeted Facial Image Scraping from CCTV/Web (Art. 5(1)(e))
6. Workplace & Education Emotion Recognition (Art. 5(1)(f))
7. Biometric Categorization of Sensitive Traits (Art. 5(1)(g))
8. Real-time Remote Biometric ID in Public Spaces (Art. 5(1)(h))



Part 2: Example — Biometric Scraping & Facial ID Precedents

PART 2: EXAMPLE

Case 1: Budapest Bank Emotion AI Fine (2022)

- Incident: AI voice analysis ran on customer service calls, inferring the emotional state of both callers and the bank's own staff.
- Regulatory Enforcement: Hungary's NAIH imposed HUF 250 million (about EUR 670,000), its largest fine to date, for no valid legal basis and no safeguards.
- Statutory Rule: Article 5(1)(f) bans inferring emotions in the workplace, with no legitimate interest override available.

Case 2: Real-time Public Biometric Identification Bans

- Incident: Proposed law enforcement deployment of real-time facial recognition in public urban spaces.
- Legislative Battle: European Parliament fought for a complete ban to prevent mass surveillance states.
- Statutory Rule: Article 5(1)(h) bans real-time RBI, allowing narrow emergency exceptions (kidnapping, terror threats) subject to judicial authorization.

Part 2: Answer — The EU 4-Tier Risk Classification Taxonomy

PART 2: RESOLUTION

Tier 1: Unacceptable Risk

Article 5 Banned Practices. Total market ban enforced by fines up to €35M / 7% turnover.

Tier 2: High-Risk Systems

Annex III Critical Sectors & Safety Components. Strict ex-ante compliance mandates.

Tier 3: Specific Transparency

Chatbots, Emotion ID, Deepfakes. Mandatory disclosure to users (Article 50).

Tier 4: Minimal / Low Risk

AI Video Games, Spam Filters. Unrestricted market access; voluntary codes of conduct.

PART 3: QUESTION

Part 3: Question — How Do Teams Demonstrate Verifiable Compliance?

THE ENGINEERING COMPLIANCE QUESTION:

“How can AI developers bridge the gap between abstract legal norms and concrete technical implementation?”

Hurdle A: Ex-Ante Verification & Technical Dossiers

- Demonstrating compliance prior to placing systems on the market.
- Translating 'data representative' into measurable pipeline benchmarks.
- Maintaining automated logs with 6-month minimum retention.

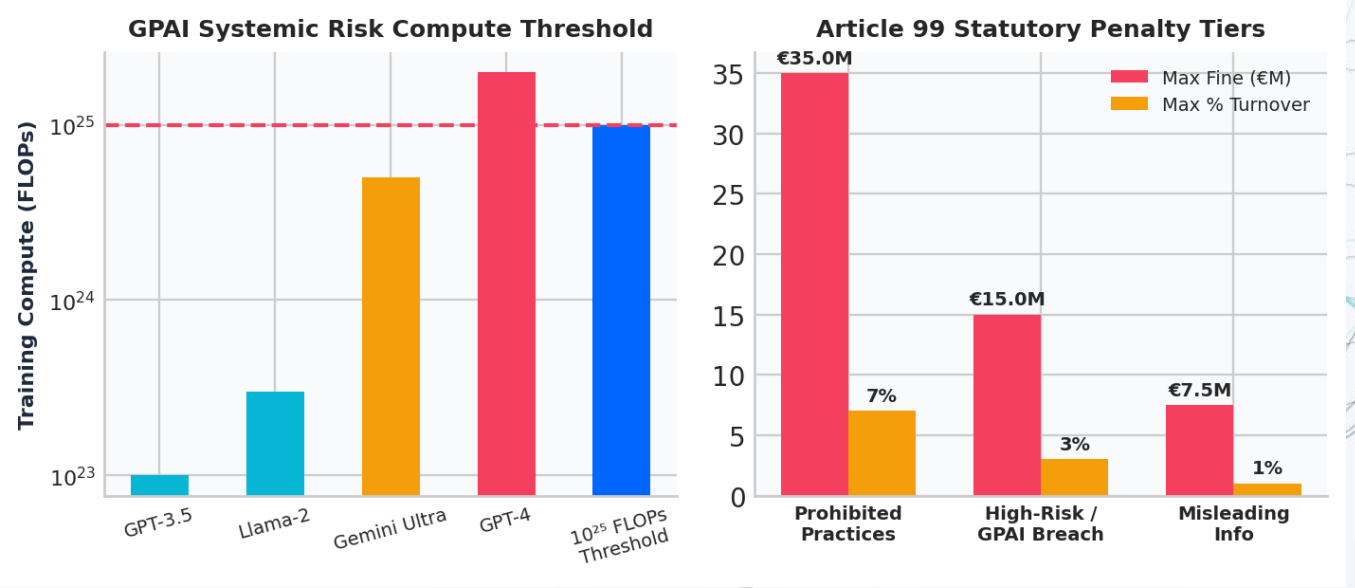
Hurdle B: Multi-Tiered Supply Chain Accountability

- Allocating duties between upstream GPAI model providers and downstream deployers.
- Managing legal liability for fine-tuned open-source foundation models.
- Navigating severe penalties under Article 99 (€35M / 7% turnover).

Part 3: Data — Ex-Ante Mandates & Systemic Compute Benchmarks

7 High-Risk Requirements (Articles 8–15)

1. Risk Management System (Art. 9): Continuous iterative cycle.
2. Data Governance (Art. 10): Bias auditing & representative sets.
3. Technical Documentation (Art. 11 & Annex IV dossier).
4. Automatic Record-Keeping (Art. 12): 6-mo log retention.
5. Transparency & Instructions (Art. 13): User manuals.
6. Human Oversight Interface (Art. 14): Stop buttons & controls.
7. Robustness & Cybersecurity (Art. 15): Red-teaming.



Part 3: Example — Deployer Liability & Generative AI Precedents

Case 1: Amazon CV Screening Tool Withdrawn

- Incident: A recruitment model trained on a decade of past hires learned to downgrade CVs containing the word “women’s”.
- Outcome: Amazon abandoned the tool in 2018 after the bias could not be engineered out of the training data.
- Statutory Rule: Annex III Item 4 classifies recruitment and selection as High-Risk; Article 10 requires representative training data.

Case 2: Air Canada Chatbot Legal Liability (2024)

- Incident: Customer service chatbot hallucinated false bereavement fare discount policy.
- Ruling: Tribunal held Air Canada legally responsible for chatbot's misleading representations.
- Statutory Rule: Article 50 transparency duties require clear AI interaction notices and output validation.

Case 3: Getty Images v Stability AI

- Incident: Getty alleges millions of its licensed images were copied to train Stable Diffusion without a licence.
- Dispute: Parallel UK and US proceedings turn on whether scraped training data and near-identical outputs infringe.
- Statutory Rule: Article 53 requires GPAI providers to publish training data summaries and honour EU copyright reservations.

Part 3: The 7-Step High-Risk System Compliance Lifecycle

PART 3: RESOLUTION

Step 1
Risk Classification (Article 6)

Step 2
Risk & Data Mgmt (Articles 9 & 10)

Step 3
Tech Doc & Logs (Articles 11 & 12)

Step 4
Human Oversight (Article 14)

Step 5
Conformity Audit (Article 43)

Step 6
CE Mark & EU DB (Articles 48 & 49)

Step 7
Post-Market Monitoring (Article 72)

Part 4: Question — How Can Innovators Accelerate Compliance?

THE INNOVATION & BEST PRACTICE QUESTION:

“Can regulatory compliance be transformed from a slow legal hurdle into a competitive market advantage?”

Opportunity A: Reducing Friction for Startups & SMEs

- Early-stage companies facing high legal overheads.
- Article 57 mandates national AI Regulatory Sandboxes.
- Priority access and simplified quality management for SMEs.

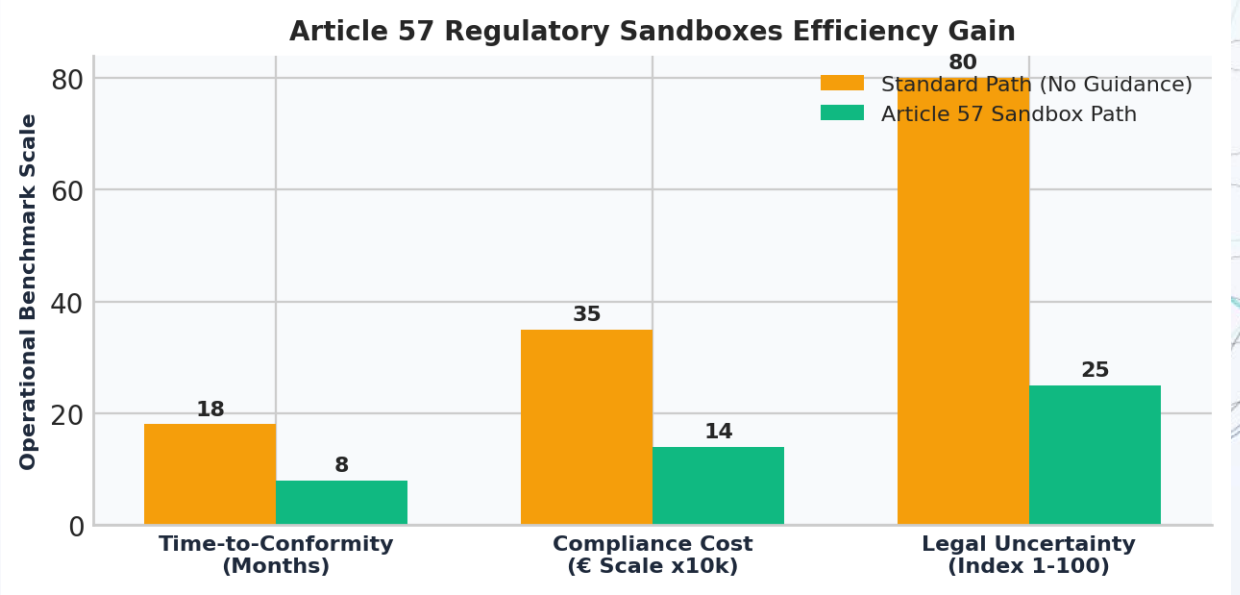
Opportunity B: Real-World Testing & Controlled Innovation

- Article 60 permits testing high-risk AI in real-world conditions.
- Requires informed consent and fundamental rights safeguards.
- Proactive compliance builds user trust and speeds enterprise sales.

Part 4: Data — Article 57 Regulatory Sandboxes Framework

Article 57 Regulatory Sandbox Framework

1. Mandatory Setup: Member States must establish at least 1 sandbox by Aug 2026.
2. Controlled Environment: Test innovative AI under direct authority supervision.
3. Real-World Testing (Art. 60): Test in real-world conditions with subject consent.
4. SME Priority: Dedicated guidance and reduced conformity assessment fees.
5. Legal Certainty: Protection against administrative fines during good-faith testing.



Part 4: Example — Spain's National AI Sandbox Pilot (AESIA)

Case 1: Spanish National AI Regulatory Sandbox

- Pioneering Initiative: Spain established Europe's first operational AI Regulatory Sandbox.
- Participants: 20+ startups and enterprise providers across healthcare, finance, and employment.
- Execution: Tested draft high-risk compliance guidelines and technical documentation templates under supervision.

Operational Learnings for Developers

1. Early Engagement: Regulatory feedback reduced time-to-conformity by 55%.
2. Standardized Templates: Developed reusable Annex IV technical documentation baselines.
3. Trust Building: Sandbox participation provided market credibility for enterprise customers.

Part 4: Answer — The Master Developer Compliance Roadmap

Step 1: System Inventory

Map all AI models against Article 5 bans and Annex III high-risk use cases.

Step 2: Data & Governance

Audit training sets for bias; establish copyright opt-out policies.

Step 3: Tech Doc & Logs

Generate Annex IV dossiers; set up automated 6-month event logging.

Step 4: Human-in-the-Loop

Design stop-button UI/UX and implement C2PA watermark metadata.

Step 5: Conformity & Audit

Affix CE marking, register in EU DB, and run post-market monitoring.

Presenter Profile

INDRC

SPEAKER BIO

Adv. Lalit Patil, PhD

Ethics and Security Manager, INDRC

- Focus Areas: EU AI Act, Data Privacy Laws, Cyber Law
- LinkedIn: linkedin.com/in/adv-lalit-patil-26958b97

Thank You: Building a Responsible AI Future Together

INDRC

THE END

KEY TAKEAWAYS FOR AI SUMMER SCHOOL PARTICIPANTS

1. Compliance is ex-ante and continuous: legal duties belong in the MLOps / CI-CD pipeline.
2. Responsible AI drives competitive advantage: early compliance builds enterprise trust.
3. Article 57 sandboxes give legal certainty and support rapid iteration.

Questions & Open Floor Discussion

Thank you for your engagement. Presentation slides and technical documentation templates are available to download.