Understanding - EU AI ACT 2024

EU AI ACT 2024 - Oerview

General Principles

- **Risk Categorization**
- **Reference:** Article 6 – Classification of AI systems.
- High-risk systems include those impacting education per Annex III.
- **Transparency**
- **Reference:** Article 52 – Transparency obligations for AI systems.
- **Accountability**
- **Reference:** Article 23 – Governance and accountability in high-risk systems.
- **Human Oversight**
- **Reference:** Article 14 – Human oversight requirements for AI systems.
- **Ethical Framework**
- **Reference:** Article 9 – Risk management system requirements.

Data Handling

- **Data Privacy**
- **Reference:** Article 10 – Quality of datasets, aligned with GDPR (General Data Protection Regulation).
- **Bias Mitigation**
- **Reference:** Article 10 – Avoiding biases in datasets.
- **Data Security**
- **Reference:** Article 15 – Cybersecurity requirements for AI systems.
- **Data Transparency**
- **Reference:** Article 13 – Documentation requirements for high-risk systems.
- **Consent**
- **Reference:** Article 52 – Consent and communication obligations for users.

AI-Driven Learning Tools

- **Content Accuracy**
- **Reference:** Article 10 – Dataset quality assurance.
- **Personalization**
- **Reference:** Article 14 – Aligning personalization with ethical oversight.
- **Feedback Mechanism**
- **Reference:** Article 54 – Reporting and feedback mechanisms for AI systems.

- **Inclusivity**
- **Reference:** Article 10 – Diverse datasets to ensure inclusivity.
- **Cultural Sensitivity**
- **Reference:** Article 9 – Risk mitigation strategies, including cultural sensitivity.

Grading and Assessments

- **Fairness**
- **Reference:** Article 7 – Prohibitions of certain AI practices (unfair grading systems).
- **Explainability**
- **Reference:** Article 14 – Human oversight to ensure explainability.
- **Error Correction**
- **Reference:** Article 56 – Error correction and liability mechanisms.
- **No Sole Decision-Making**
- **Reference:** Article 14 – Human oversight in decision-making processes.
- **Accuracy Validation**
- **Reference:** Article 10 – Continuous testing for dataset and model accuracy.

Student Engagement

- **Interaction Design**
- **Reference:** Article 14 – Ensuring systems are designed for responsible usage.
- **Feedback Personalization**
- **Reference:** Article 9 – Personalization with fairness considerations.
- **Privacy in Chatbots**
- **Reference:** Article 52 – Transparency in AI communication tools.
- **Monitoring Usage**
- **Reference:** Article 15 – Usage monitoring and cybersecurity protocols.
- **Emotional AI Limitations**
- **Reference:** Article 5 – Prohibition of harmful or manipulative AI systems.

Training for Educators

- **AI Literacy**
- **Reference:** Article 9 – Training and awareness programs for stakeholders.
- **Bias Awareness**
- **Reference:** Article 10 – Educator training on dataset biases.
- **Decision Oversight**
- **Reference:** Article 14 – Human decision-making training requirements.
- **Ethical Use Training**

- **Reference:** Article 9 – Awareness programs on ethical AI use.
- **Policy Awareness**
- **Reference:** Article 23 – Internal policies for AI governance in organizations.

Procurement and Deployment

- **Supplier Compliance**
- **Reference:** Article 24 – Supply chain management obligations.
- **Documentation**
- **Reference:** Article 13 – Comprehensive documentation of AI system operations.
- **Risk Assessment**
- **Reference:** Article 9 – Risk management plan for deployment.
- **Third-Party Audits**
- **Reference:** Article 20 – Conformity assessments by third parties.
- **Regular Updates**
- **Reference:** Article 13 – Documentation to reflect system updates.

Special Needs and Accessibility

- **Accessibility Features**
- **Reference:** Article 14 – Inclusion of human oversight for accessibility.
- **Support for Disabilities**
- **Reference:** Annex III – High-risk systems, including those designed for disabilities.
- **Language Support**
- **Reference:** Article 10 – Dataset diversity, including multilingual data.
- **Customizable Interfaces**
- **Reference:** Article 9 – Designing customizable features for inclusivity.
- **Assistive AI Validation**
- **Reference:** Article 20 – Validation and testing for assistive technologies.

Long-Term Impact

- **Future-Proofing**
- **Reference:** Article 13 – Regular updates and adaptable documentation.
- **Sustainability**
- **Reference:** Article 9 – Environmental considerations in AI usage.
- **Cost-Benefit Analysis**
- **Reference:** Article 20 – Economic assessment during conformity checks.
- **Scalability**
- **Reference:** Article 13 – Design requirements for scalability.

- **Continuous Improvement**
- **Reference:** Article 54 – Reporting and adapting AI systems.

Stakeholder Engagement

- **Parental Involvement**
- **Reference:** Article 52 – Transparency obligations to inform all stakeholders.
- **Student Participation**
- **Reference:** Article 54 – Mechanisms for user feedback.
- **Community Outreach**
- **Reference:** Article 23 – Institutional governance including community input.
- **Cross-Institution Collaboration**
- **Reference:** Article 20 – Shared practices through audits and testing.
- **Regulatory Liaison**
- **Reference:** Article 62 – Coordination with regulatory authorities.

Article 6 – Classification of AI Systems

1. Four Risk Levels

AI systems are classified into **four categories** based on their risk potential:

- **Prohibited AI Systems:** AI practices that pose unacceptable risks and are banned.
  - Examples: Subliminal manipulation, exploitation of vulnerabilities, and social scoring systems by public authorities.
- **High-Risk AI Systems:** Systems that significantly impact individuals' safety or fundamental rights.
  - Examples: AI used in biometric identification, critical infrastructure, education, employment, credit scoring, or healthcare.
- **Limited-Risk AI Systems:** Systems requiring transparency obligations but not as tightly regulated as high-risk systems.
  - Examples: Chatbots, recommendation systems, and virtual assistants.
- **Minimal-Risk AI Systems:** Systems with negligible risk, which are largely unregulated.
  - Examples: Entertainment AI, spam filters, and AI-powered games.

---

2. Criteria for Classification

The classification process considers:

- **Sector of Application:** The domain where the AI is deployed (e.g., education, healthcare).
- **Impact on Rights and Safety:** How the AI affects individuals' privacy, safety, or fundamental rights.
- **Severity of Harm:** The potential damage caused by incorrect or biased outcomes.
- **Autonomy of AI Decision-Making:** The level of human involvement or oversight in the AI's decisions.

---

3. High-Risk Systems in Education

Educational AI systems are explicitly listed under **Annex III** of the EU AI Act as high-risk if they:

- Determine **student access to education** (e.g., AI used in admissions).
- Influence **learning outcomes** (e.g., AI-powered grading systems).
- Assess **skills or competencies** that significantly affect career prospects.

These systems must comply with strict regulations, including documentation, risk management, and transparency requirements.

---

4. Obligations for High-Risk Systems

For AI systems classified as high-risk, the following obligations apply:

- **Conformity Assessments:** Systems must pass pre-deployment evaluations for compliance.
- **Risk Management:** Continuous risk assessment throughout the system's lifecycle.
- **Data Requirements:** High-quality, representative, and bias-free datasets.
- **Human Oversight:** Mechanisms to ensure human intervention when needed.
- **Monitoring and Reporting:** Continuous monitoring of performance and reporting of incidents.

Implications for Stakeholders

AI Developers

- Must evaluate whether their system falls under high-risk categories.
- Implement safeguards like robust testing and documentation.

Educational Institutions

- Ensure AI tools for admissions, grading, or skill assessment meet high-risk criteria.
- Conduct regular audits to verify compliance with the EU AI Act.

Regulators

- Monitor the deployment of AI systems in high-risk areas, especially those influencing fundamental rights.
- Enforce penalties for non-compliance.
