Presentation 02

Ms. Yunkyoung Kim

Senior Research Associate, Center for International Development
Korea Development Institute (KDI)

Pillars on AI Regulation

October 3, 2024 | Yunkyoung Kim

Progression of AI Regulation Initiatives

Year	Framework	Description
May 2019	**OECD AI Principles**	The first international AI principles addressing transparency, accountability, inclusivity, fairness, robustness, and privacy
Feb 2020	**OECD AI Policy Observatory**	A platform launched to track and analyze global AI policies and regulatory trends
2021	**NIST AI Risk Management**	Initial discussions on developing a framework to manage AI risks and ensure reliable AI systems
Apr 2021	**EU AI Act Proposal**	A proposed regulatory framework based on categorizing AI risks to guide AI governance in the EU
Nov 2021	**UNESCO AI Ethics Recommendation**	Guidelines proposed by UNESCO to address ethical concerns in AI development and deployment
Jan 2023	**NIST AI Risk Management Framework 1.0**	The finalized version of NIST's framework for managing AI risks and improving system trustworthiness
Mar 2023	**UK AI Regulation: A pro-innovation approach**	The UK government's proposed framework aimed at promoting innovation while applying existing regulations to AI
May 2023	**G7 Hiroshima AI Principles**	Principles adopted by G7 nations to promote international collaboration on AI transparency, safety, accountability, and innovation
May 2024	**EU AI Act Adoption**	The European Council adopted the EU AI Act, moving towards the implementation of a comprehensive AI regulatory law

AI-related Issues

Job Displacement	Privacy Infringement	Bias and Discrimination	Safety Risks	Intellectual Property Disputes	Unfair Competition
Security Threats	Hate Speech	Trade Barriers	AI-Driven Misconduct	Identity Thefts	Unequal Access to Technology
Political Manipulation	Economic Inequality	Misinformation and Fake News	Misuse in Weapons	Environmental Impact	Algorithmic Transparency

Key Issues in AI: Direct vs. Indirect Areas of AI Impact

Issues Directly Related to AI Technology and Systems			Issues Indirectly Related to AI Technology and Systems			
Data Collection and Processing	**Ensuring Safety**	**Ethical Use**	**Economic Impact**	**Legal and Regulatory Impact**	**Social Impact**	**Environmental Impact**
• Personal Data Breaches • Security Threats • Unauthorized Data Collection • Data Quality and Bias	• System Safety • Unpredictable Behavior and Errors • Public Safety, Cybersecurity, and National Security	• Fairness, Discrimination • Human Rights Protection • Privacy Infringement • Diversity • Transparency • Social Responsibility	• Market Monopoly and Unfair Competition • Job Displacement and Economic Inequality • Innovation and Industrial Transformation	• Conflicts with Existing Laws and Regulations • Intellectual Property and Personal Rights • Ownership of AI-Generated Content	• AI Inclusivity, Accessibility, and Literacy • Deepening Education and Technology Gaps • Psychological Impact and Mental Health • AI-Driven Crime and Spread of Fake News	• Resource Concentration, including Energy • Intensification of Climate Change and Environmental Pollution • Climate Change Mitigation through Innovation

- ***Technology-specific Issues*:** focuses on technology-specific issues, requiring new regulatory approaches that include technical adjustment and solutions.
- ***Broader Societal Impacts*:** addresses indirect issues arising from the broader societal impacts of AI, necessitating adaptations to existing frameworks.
- ***Comprehensive View of AI Regulation*:** shows the interconnection between technology-specific issues and broader societal impacts, while considering diverse regulatory approaches for each.

AI Regulation: Converging on Data, Safety, and Ethics

	Data	Safety	Ethics	
Major AI Regulatory Frameworks	**Data Privacy, Data Quality and Bias**	**System Stability, Reliability, Robustness, Security**	**Fairness, Transparency, explainability, Accountability, Bias Mitigation**	**Other**
OECD AI Principles	–	Robustness, safety	Human-Centered Values, Fairness, Transparency, Accountability	Inclusive Growth, Well-being Sustainable Development
EU Ethics Guidelines For Trustworthy AI	Privacy and Data Governance	Technical Robustness and Safety	Human agency and Oversight, Transparency, Diversity, Non-discrimination, Fairness	Social and Environmental Well-being
EU AI Act, High-Risk AI System Requirements	Data Governance	Safety, Accuracy, Robustness and Cybersecurity	Transparency and Explainability	Technical Documentation, Record-keeping, Human Oversight
USA: Blueprint for an AI Bill of Rights	Data privacy	Safe and Effective Systems	Algorithmic Discrimination Protections	Notice and Explanation, Human Alternatives
USA: NIST AI Risk Management Framework	Data Quality	Validity, Reliability, Safe, Security, Resilience	Accountability, Transparency, Privacy-Enhanced, Fair–with Harmful Bias Managed	–
UK: AI Regulation: pro-innovation approach	–	Safety, Security, Robustness,	Transparency, Explainability, Fairness, Accountability	Governance, Contestability, Redress
Canada: Directive on Automated Decision-Making	Quality Assurance	Security	Algorithmic Impact Assessment, Transparency	Recourse, Reporting
China: Ethical Norms for the Next Generation AI	Data Quality	Security, Reliability, Predictability	Transparency, Accountability, Fairness, Justice, Privacy	Human Welfare
Korea: AI Ethics Guidelines	Data Management	Prohibition of Harm, Safety	Human Rights Protection, Privacy Protection, Public Interest, Respect for Diversity, Accountability, Transparency	Solidarity

Data Collection and Processing

Key Issues / Events

- Privacy breaches, security threats, unauthorized data collection
- Cambridge Analytica / GDPR (2018), CCPA (2020)

Regulatory and Management Areas

- Privacy protection, data management and security, data ownership and access rights, data transparency
- **Application Examples**: Privacy protection, data minimization, data anonymization, data sovereignty, Privacy-Enhancing Technology (PET)

Major Regulatory Frameworks

- Legal frameworks, regulatory legislation, data protection laws
- **Examples**: GDPR (EU), CCPA (US), PIPEDA (Canada)

Regulatory Consideration

- **Technological**: Data quality, suitability, accuracy, privacy protection assurance
- **Social**: Data ownership, transparency, user consent and rights
- **Cultural**: Cultural differences in data privacy and security perceptions

Ensuring Safety

Key Issues / Events

- System vulnerabilities, cyberattacks, data corruption or loss
- Autonomous vehicle accident (2016), financial AI system error (*Knight Capital, Flash Crash* cases)

Regulatory and Management Areas

- System reliability, predictability, risk assessment and management, standards compliance, and validation
- **Application Examples**: Safety and security of Autonomous vehicles, medical AI systems, and financial AI systems

Major Regulatory Frameworks

- Technical standards, safety frameworks, regulatory certification systems
- **Examples**: ISO/IEC Standards, NIST RMF, EU AI Act

Regulatory Consideration

- **Technological**: System security, stability, reliability, robustness
- **Social**: System predictability, risk management, user protection, public safety
- **Cultural**: Cultural differences in safety perceptions and risk tolerance levels

Ethical Use

Key Issues / Events

- Societal acceptance of AI technology, human rights, bias, discrimination
- Recruitment AI systems (2018), Risk Assessment Tools in Criminal Justice (2016)

Regulatory and Management Areas

- Discriminatory algorithms and biased data processing, protection of individual privacy
- **Application Examples**: Fairness and transparency in recruitment AI systems, credit scoring AI, human rights protection in facial recognition AI

Major Regulatory Frameworks

- Ethical guidelines, international principles, recommendations
- **Examples**: Asilomar AI Principles, EU Commission Trustworthy AI Guidelines, UNESCO AI Ethics Guidelines, OECD AI Principles

Regulatory Consideration

- **Technological**: Algorithm fairness, bias mitigation, explainability
- **Social**: Social responsibility, human rights protection, non-discrimination
- **Cultural**: Cultural differences in ethics, perceptions of fairness and justice

Three Pillars of AI Regulation

Interconnected Pillars with Divergent Priorities

- **Interconnected**: collaborating to achieve the common goal of safe and trustworthy AI.
- **Priorities**: may vary by issue and can sometimes be contested.

Foundations of Effective AI Regulation

- Effective AI regulation requires an understanding of associated risks and a structured approach, based on key pillars.
- This comprehensive approach serve as the basis for crafting effective regulation and navigating the complex AI landscape.

Leveraging Pillars for Broader Societal Impact

- Regulatory frameworks for broader societal impacts of AI should be grounded in these key pillars.
- By starting with these pillars, we can expand them as tools tackle broader societal issues.

Harms and Risks of AI

Harms of AI		Risks of AI		
Harms	**Key Issues**	**Data Collection & Processing**	**Ensuring Safety**	**Ethical Use**
Social Harm	Privacy infringement	●	○	◎
	Unfair competition	●	○	◎
	Behavioral manipulation	◎	○	●
Economic Harm	Excessive/biased automation	○	◎	●
	Loss of human judgement	○	◎	●
	Excessive monitoring	◎	○	●
Political Harm	Social media bias	●	○	◎
	Data and information control	●	○	◎
	Erosion of political discourse	◎	○	●

Note: ●>◎>○ - Listed according to their relevance between main issues of harm and key areas in AI regulation.

Towards International Collaboration

Data

- **Continuous Sharing**: regularly share strategies for data management and privacy protection.
- **Innovation**: where possible, extend the sharing of public data to foster innovation.

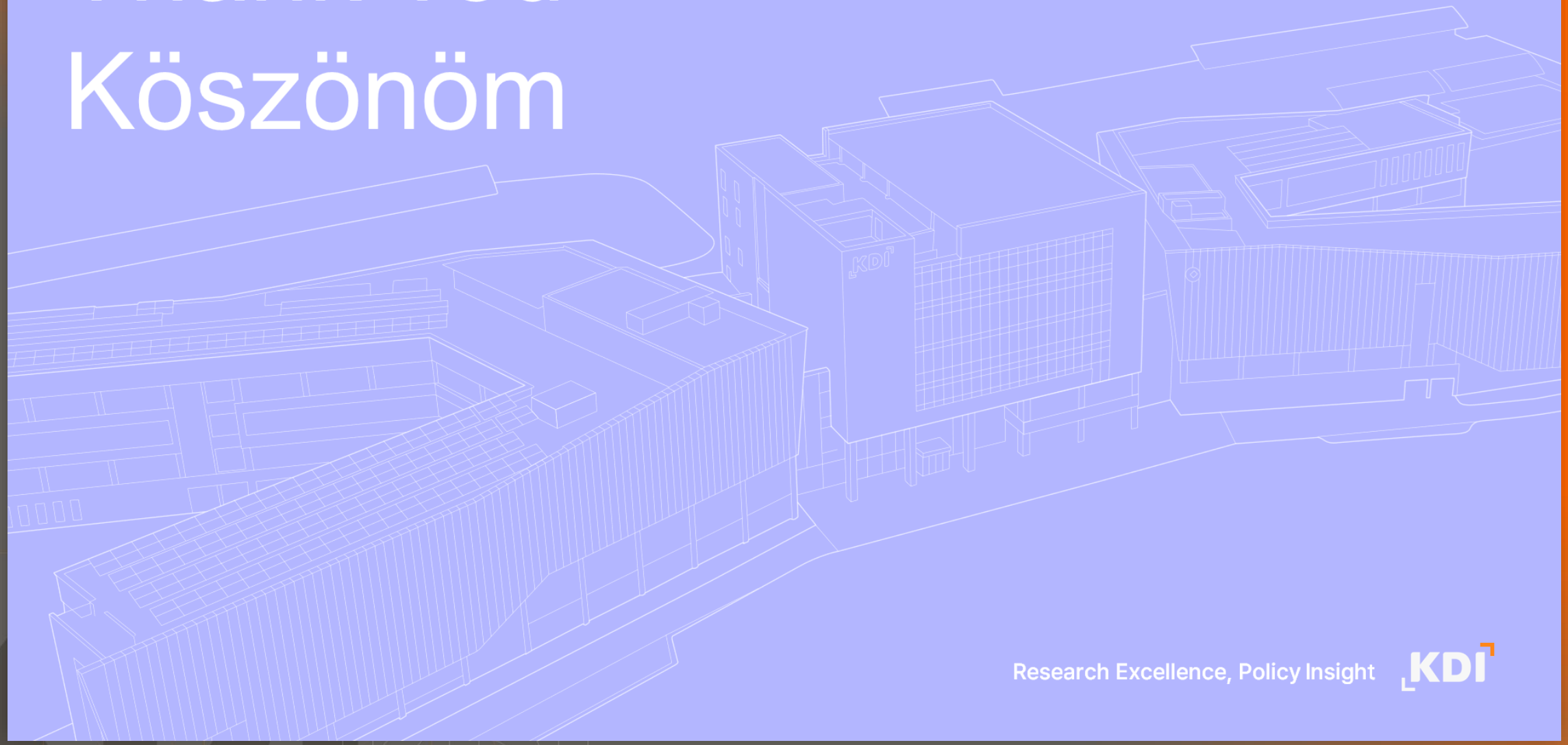
Safety

- **The AI Safety Summit**: Initiated active discussions on safety by emphasizing the development of AI safety standards.
- **Ongoing Enhancement**: involves continuous monitoring, aligning standards, and sharing best practices, and guidelines to improve safety measures.

Ethics

- **AI Ethics Challenges**: AI ethics regulations are in the early stages, facing challenges in integrating ethical considerations into practical implementation strategies.
- **Essential Collaboration**: sharing knowledge and information, and developing strategies using case studies on ethical issues and practical applications.

Thank You
Köszönöm



Research Excellence, Policy Insight