United Nations University
Institute in Macau

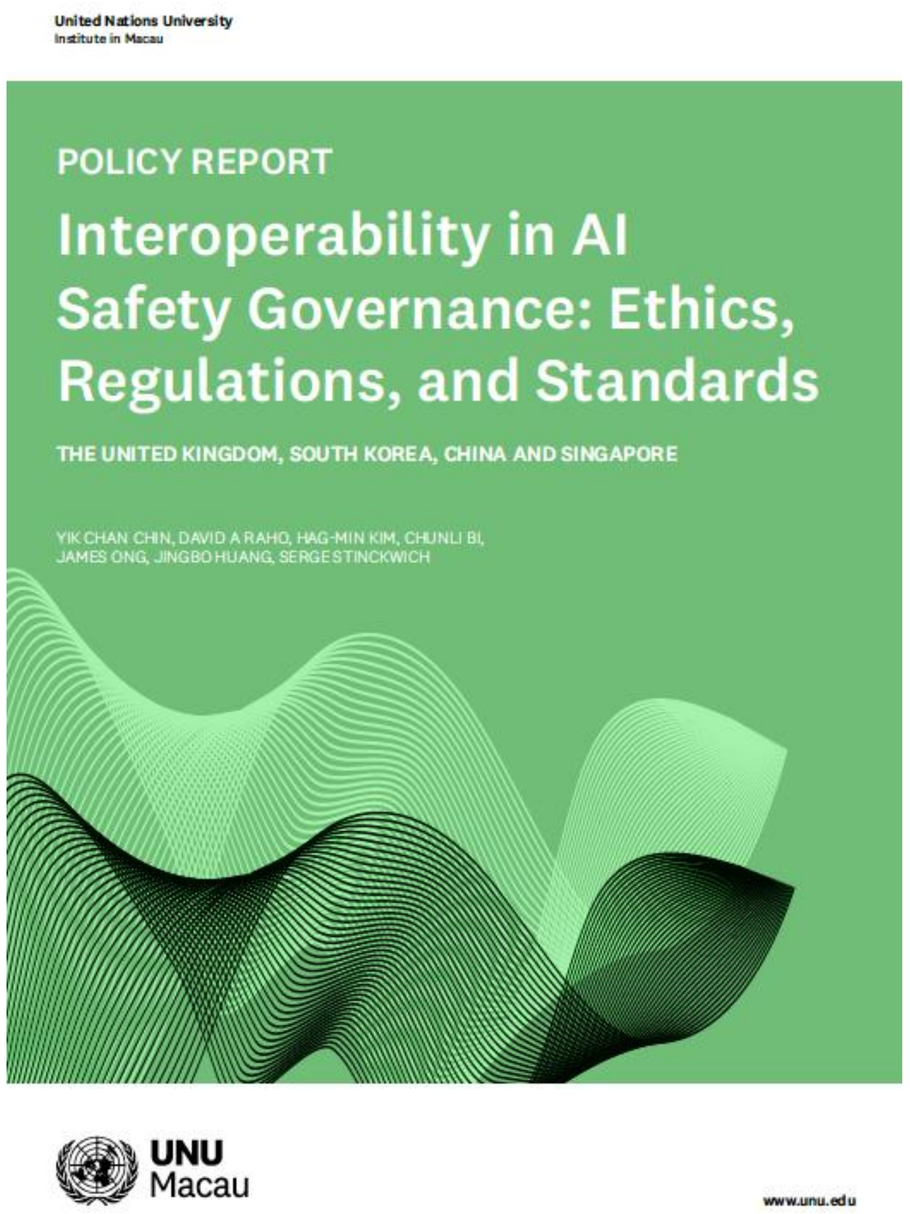
Interoperability in AI Safety Governance: Ethics, Regulations, and Standards

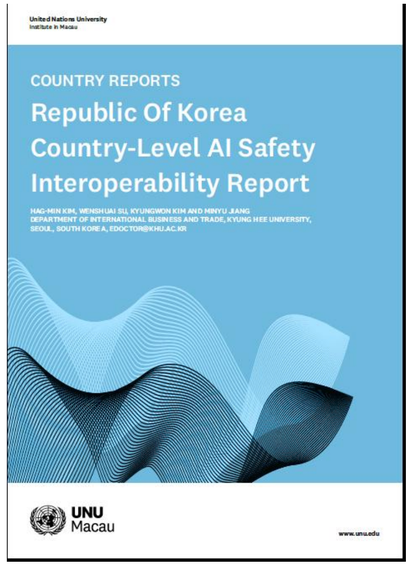
THE UNITED KINGDOM, SOUTH KOREA, CHINA AND SINGAPORE

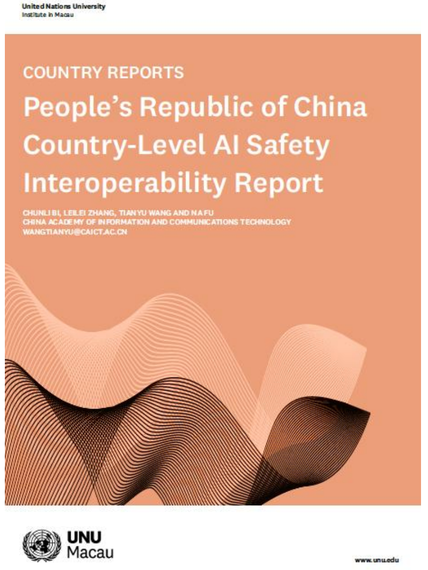
YIK CHAN CHIN, DAVID A RAHO, HAG-MIN KIM, CHUNLI BI, JAMES ONG, JINGBO HUANG, SERGE STINCKWICH

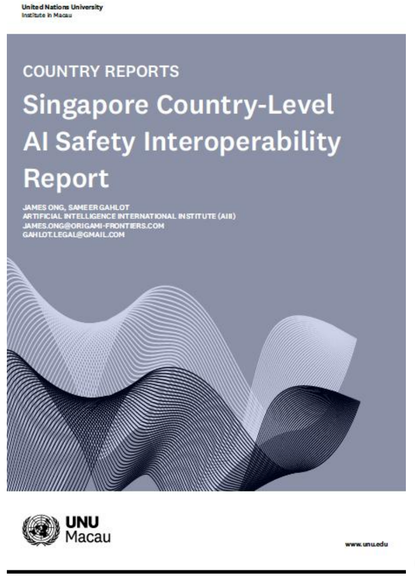
yik-chan.chin@bnu.edu.cn

UNU Macau









UNU Macau

https://collections.unu.edu/eserv/UNU:10363/Interoperability_in_AI_Safety_Governance.pdf

Purpose of the Report

- This policy report draws on four country studies from **China, South Korea, Singapore, and the United Kingdom** to identify effective tools and key barriers to interoperability in AI safety governance, develops policy recommendations to support the creation of interoperability mechanisms aligned with the Global Digital Compact (GDC) and UN resolutions on AI safety governance.
- 2024 two UN General Assembly resolutions on AI: governance measures must be interoperable, flexible, adaptable, inclusive etc.
- 2024 GDC emphasises the importance of interoperability in AI governance across its various scopes: Coordination, interoperability, and compatibility of emerging AI governance frameworks (Objective 5 of the GDC) are promoted through
  - Establishing the International Scientific Panel on AI and the Global Dialogues on AI Governance.
  - Sharing best practices and promoting common understanding in AI. (ethics)
  - Encouraging transparency, accountability, and strong human oversight of AI systems in line with international law. (regulations)
  - Encouraging standards development organisations to collaborate on interoperable AI standards that uphold safety, reliability, sustainability, and human rights. (standards)
  - Establishing international partnerships to develop education and training programmes, increase access to open AI models and systems, share training data and computing resources, and support AI model training and development. (Int'l collaborations)
  - Addressing local needs, fostering cross-regional partnerships, and connecting them globally to ensure AI interoperability frameworks are inclusive, adaptable, and capable of tackling local challenges. (local needs)
  - Establishing a dedicated working group on data governance under the Commission on Science and Technology for Development.

UNU Macau

Purpose of the Report

- Interoperability is a central goal of AI governance, vital for reducing risks, fostering innovation, enhancing competitiveness, promoting standardization, and building public trust.

- Focusing on three high-stakes domains-**autonomous vehicles, education, and cross-border data flows**
  - Salient risks: threats to the right to privacy, the right to life, and the right to equitable access to knowledge and digital literacy, data security and vehicle cybersecurity.
  - the UN's AI Safety framework: safe, secure, and trustworthy AI systems, uphold human rights, promote sustainable development, and be governed by ethical principles throughout their lifecycle

UNU Macau

Define the Terms

- Safety of an AI system: the understanding, prevention, mitigation, and management of potential harms arising from the design, development, and deployment of AI systems, ensuring that AI technologies protect human well-being throughout their lifecycle.
  - These safety harms may be deliberate or accidental, and can affect individuals, groups, organizations, nations, or even global systems, taking various forms such as physical, psychological, or economic impacts, Examples: algorithmic bias, privacy leakage, misinformation and deepfakes, unreliable decision-making.

- AI safety governance: encompasses frameworks, policies, and operational practices that ensure AI is developed, deployed, and maintained in a safe, reliable, and ethical way, reducing risks and avoiding harm to individuals and society (Jobin et al., 20193; Lee et al., 20214; Tabassi, 20235; OECD, 2019/20246).

- Interoperability: the ability of different systems, tools, and components to work together seamlessly both technically through enabling data sharing and normatively through aligning laws and standards (PNAI 2023& 2024; Zeng, 2019; Berg, 2024; Onikepe, 2024).
  - substantive measures such as international norms, shared protocols, interfaces, and data models, enabling communication, data exchange, standardizations etc

- Interoperability of AI safety governance: essential substantive methods that enable two or more different jurisdictions to collaborate in order to support a common understanding, interpretation, and implementation of transborder AI safety governance.
  - Greater interoperability can reduce risks, foster innovation, enhance competitiveness, promote standardization, and build public trust.

UNU Macau

Define the Terms

Three key aspects of interoperability in AI safety governance: ethical, legal and technical interoperability.

Functions of Ethical interoperability	The ability of institutions, systems, or actors to collaborate across different moral frameworks to support the development of AI regulations and technical standards, as well as international cooperation.
Functions of Legal interoperability	Involves the coordination of regulatory frameworks and establishing international cooperative mechanisms. The development of legal interoperability's substantive and structural dimensions as a "third way" between fragmentation and harmonisation merit increased attention.
Functions of Technical interoperability	Ensures the compatibility of AI technical standards in addressing governance issues related to technical interconnectivity, transactional interconnectivity, physical externalities, and policy externalities.

TECHNICAL INTEROPERABILITY

"The ability of two or more systems or components to exchange information and to use the information that has been exchanged", focusing on ensuring systems can communicate and work together.

- Vertical interoperability: interoperability between applications or systems on different stack levels of the stack.
- Horizontal interoperability: interactions between applications or systems in the same stack layer of the stack that provide similar or complementary functionality.

UNU Macau

Methods

- Regulatory learning: learning from domestic experiments, ideas, experiences, and insights of other jurisdictions or communities.
  - To determine the most suitable regulatory response to the challenges posed by AI, interjurisdictional learning and regulatory innovation are necessary through the exchange of ideas and experiences.
  - Learning to regulate AI effectively requires robust international cooperation and coordination
  - Stakeholders from four jurisdictions to harness local insights and promote interjurisdictional learning.
- Led and coordinated by the United Nations University (UNU), the initiative pioneers a new approach to transnational policy-making
  - Defining shared challenges, and developing collaborative strategies.
  - Generate practical solutions to AI safety governance's complex and pressing policy challenges, fostering a globally informed yet locally grounded regulatory ecosystem.
- A framework to compare AI safety governance across four jurisdictions:
  - **Seven Elements: Objectives; Principles and values; Governance approach (Top-Down vs. Bottom-Up); Binding nature; Level of integrations; Regulator; Components.**

UNU Macau

Table 3: Comparison of AI safety governance of four jurisdictions China, South Korea, Singapore and UK (Transborder Data Flow)

Jurisdiction	Level of Integration	Components of framework: 1) objectives; 2) ethics; 3) binding; 4) targeted regulations/frameworks; 5) technical standards; 6)regulators; 7)risks/challenges
China	China has established a structured framework for integrating its cross-border data flow governance with international norms, and is striving to achieve alignment with international standards. • **Legal and Principle Alignment:** Its core regulatory framework (anchored in the Data Security Law, Personal Information Protection Law (PIPL), and Provisions on Promoting and Regulating Cross-Border Data Flows (2024)) aligns with OECD Data Governance Principles, emphasizing "data free flow with trust" and balancing security with legitimate cross-border cooperation. It adopts global technical standards such as TLS 1.3 for encrypted transmission and ISO/IEC 27701 for privacy management, while GB/T 43697-2024 (Data Classification and Grading Rules) aligns with international data risk assessment methodologies. • **International Engagement and Mutual Recognition:** China participates in multilateral forums including the OECD Working Party on Data Governance and GPAI's Data Governance Working Group, and acts as an observer in the Global CBPR Forum to promote standard mutual recognition. China has established a comprehensive free trade relationship with Singapore and conducts overall planning and coordination of cross-border data rules. Via the "Digital Silk Road," it also provides technical assistance to developing countries to build data security frameworks, expanding global trust networks.	**1. Objectives** • Establishing a "data sovereignty with trust" system to balance national security, individual privacy, and legitimate cross-border data needs. • Promoting international cooperation via bilateral/multilateral agreements and participation in global forums. • Supporting low-risk cross-border data activities through exemption mechanisms (per 2024 Provisions on Promoting and Regulating Cross-Border Data Flows). • Advancing privacy-enhancing technologies such as federated learning and edge computing to reduce cross-border raw data transmission. **2. Binding Nature: Legally binding requirements** • Mandatory domestic storage of personal information and important data for Critical Information Infrastructure Operators (CIIOs) (Cybersecurity Law, 2017). • Compulsory data export safety assessments for transfers of "important data" or personal information of 100,000+ individuals (Data Security Law, 2021; Measures for Data Export Safety Assessment, 2022). • Mandatory filing of standard contracts for cross-border personal information transfers involving <100,000 individuals (Measures for the Administration of Personal Information Export Standard Contracts, 2023). • Annual compliance audits for enterprises engaged in cross-border data transfers. • Non-binding flexibility: Exemptions for low-risk activities per Provisions on Promoting and Regulating Cross-Border Data Flows (2024). **3. Principles/Values** • "Data sovereignty with trust": Ensuring cross-border data flows uphold national security and individual rights. • "Fair and equitable governance": Discouraging excessive data localization requirements that hinder legitimate cross-border cooperation. • Data security and privacy protection: Embedding "data minimization," "anonymization," and "pre-transfer risk assessment" into cross-border data rules. • Sustainability: Promoting green data flows. **4. Technical Standards** • GB/T 35273-2020 Information Security Technology—Personal Information Security Specification: Provides technical criteria for identifying sensitive data and assessing cross-border risks. • GB/T 43697-2024 Data Security Technology—Data Classification and Grading Rules: Guides technical classification of data to determine cross-border risk levels. • Personal Information Protection Certification Implementation Rules (2022): Sets technical requirements for certification; recognized in over 10 countries via bilateral agreements. • TC260 Artificial Intelligence Safety Governance Framework (2024): Includes guidelines for auditing AI models trained on cross-border data to ensure compliance with data origin laws. • TLS 1.3 encryption: Mandated for secure cross-border data transmission. • ISO/IEC 27701 (Privacy Management): Adopted to align with global privacy technical standards.
South Korea	• Dual-track engagement working with both regulatory-driven and industry-driven models to cover more ground • Robust foundation for AI interoperability, especially in personal data transfer regulation; Strong alignment with global data standards (GDPR adequacy, new Global CBPR member, etc.); Integrated legal framework with emerging international standards on data protection; • G7's "Data free flow with trust" – cross-border innovation with preserving rigorous privacy and safety standards • Seoul Declaration for Safe, Innovative and Inclusive AI, a high-level pledge to cooperate on interoperable AI governance standards across borders. • Exploring mutual recognition of AI audits and certifications with partner countries • Mirroring GDPR's "continuity of protection" principle; • Data Protection Impact Assessment (DPIA) aligned with GDPR methodology -interoperability of risk management approaches • Contributing to ISO/IEC and ITU working groups on AI and data standards, OECD on AI system risk classification , ISO/IEC JTC 1/SC 42 on AI	**1. Objectives** • Stablish AI safety and security-by-design: Safe AI deployment with robust data security – and vice versa – emphasize a holistic approach. **2. Binding Nature** Legally binding requirements: • AI Framework Act (2024): establishes obligations for "high-impact AI" systems (those affecting safety or basic rights) in critical sectors, requiring risk assessments, impact evaluations, and transparency measures. • Preventing personal data misuse abroad: protections "travel with the data" (through legal and technical bindings) Non-binding flexibility: • Preventing personal data misuse abroad: oversight "follows the data" (through cooperation and representative arrangements). **3. Principles/Values** • Structured, principle-based controls: prior assessment, documented safeguards, downstream restrictions, and individual empowerment. From pre-transfer risk assessments (to prevent unsafe or unethical data uses) to post-transfer monitoring and enforcement (to correct any harms). Active enforcement by PIPC (fines, model deletion orders). **4. Technical Standards** • MyData API for secure data portability with privacy by design, all major data controllers must implement standardized APIs. • Technical standardization of data formats and semantics. • "Secure data corridors." – Investing in secure international network links and clouds. • Mandated pre-transfer risk assessment mandated with standardized risk assessment templates and software. • AI Framework Act anticipates adopting international technical standards: it includes provisions that Korean AI assessment criteria should reference globally recognized standards. Korea's National AI Standards Council has already adopted dozens of ISO/IEC AI standards as KS (Korean Standards). **5. Targeted Legislation/Framework** Foundational laws: • Personal Information Protection Act (PIPA): A "right to data portability"; A right to explanation of algorithmic decisions. Sector-specific regulations: • AI Framework Act (2024). Data-specific rules: • PIPC and Korea Internet & Security Agency (KISA) launched domestic CBPR certification system. • In 2025, PIPC announced plans to adopt its own "whitelist" of countries with equivalent protection, starting with the EU and also evaluating others like the UK, U.S., and Japan for potential adequacy determinations.

INTEROPERABILITY OVERVIEW

The major integration measures each jurisdiction aligned with global or regional AI governance frameworks are identified in three tables (Table 6, Table 7 and Table).

Jurisdiction	Level of Integration
China	• China's AV governance achieves strong alignment with international technical standards and engages in targeted bilateral cooperation, and is committed to ensuring the consistency between domestic standards and international standards. • **Technical Standard Alignment:** Key domestic standards align with global benchmarks: GB/T 40429-2021 (AV Grading) adopts UNECE WP.29's automation classification; the GB/T 34590 series is equivalent to ISO 26262 (functional safety); and GB 44495-2024 (Vehicle Cybersecurity) references ISO/SAE 21434. It also aligns with UNECE WP.29 regulations on automated driving, ensuring compatibility with global AV terminology and safety validation frameworks. • **Bilateral/Multilateral Cooperation:** China and Germany have signed the Joint Statement of Intent on Cooperation in the Field of Autonomous and Connected Driving, and will jointly develop vehicle-to-everything (V2X) technology. Domestically, 34 AV pilot zones (e.g., Beijing) adopt practices (e.g., unified operation data platforms, "black boxes" for event logging) consistent with international AV safety monitoring norms.
South Korea	• **Technical Standards Alignment:** Actively engages in joint V2X/C-ITS R&D and cross-border pilots via MoUs and standardization working groups. Actively participates in UNECE WP.29 (incl. UN R155 cybersecurity, UN R156 software updates) and ISO/TC 204, aligning national rules with evolving UN/ISO standards. • **International Alignment:** Korea adapted KMVSS to UN Regulation No. 157 (ALKS). Implemented via MOLIT Notice No. 2022-670, enhancing consistency with UNECE rules. • **National Testability & Verification:** Operates "K-City" as a multi-scenario proving ground (highway, urban, suburban, parking, community facilities). Validation platform is implemented for 5G/C-ITS integration. The vehicle–infrastructure interoperability is verified in real traffic environments.
Singapore	• Participates in the Asia-Pacific Economic Cooperation (APEC) to broaden technical coordination and support harmonized standards and regulatory approaches. • Aligns with UN regulation on Cybersecurity Management Systems that refers to standards like ISO 26262 for Functional Safety and ISO/SAE 21434 for Cyber-security of Road Vehicles.
UK	• **Interoperability by design**, enabling UK and foreign AVs to operate safely across different markets. • **Harmonise AV regulations** through the UNECE and bilateral agreements. · Based on international safety standards, such as UNECE regulations and ISO standards, by incorporating specific measures designed for autonomous functionality. Standardisation initiatives cover terminology and scenario descriptions that are essential for interoperability and safety validation. • Horizon Europe and partnership accords with countries.

Table 8: Comparison of AI safety governance's interoperability of four jurisdictions China, South Korea, Singapore and UK (Autonomous Vehicles)

UNU Macau

EFFECTIVE INTEROPERABILITY INSTRUMENTS

EFFECTIVE INTEROPERABILITY INSTRUMENTS

1) Convergency: The instruments adopted by most of the four jurisdictions in their global or regional integrations;
2) Complementarity: Different governance instruments that reinforce each other to achieve complementary performance, leading to more robust, ethical, and effective integration and oversight of AI safety (e.g., AI regulations can be complemented by liability rules to address AI harms).

Ethical interoperability

- **Promotion of shared or common terminology**: identify shared or universal terminologies capable of articulating diverse ethical principles.
- **Compatible cross-institutional and international ethical assessment and accountability mechanisms**: structured processes for disclosure and evaluation. Adopting institutions can then assess whether these characteristics align with their own ethical principles.
- **Multi-stakeholder engagement**: collaborative engagement across key stakeholders such as AI developers, ethicists, healthcare professionals, legal experts, policymakers, and community representatives, and establishing platforms to support continuous dialogue

UNU Macau

INTEROPERABILITY INSTRUMENTS

Legal interoperability:

- Harmonisation: The process of unifying law, often building on a prior approach of standardisation. Harmonisation can be applied to varying extents.
- Standardisation: A regulatory approach grounded in widely accepted principles, practices, or guidelines within a specific field. Standardisation serves as a foundational step toward eventual harmonisation, facilitating alignment across jurisdictions and sectors.
- Mutual Recognition : A principle that assesses whether regulatory measures between countries are comparable or equivalent. It reflects an agreement in which one country may relinquish a degree of regulatory independence by accepting that another nation’s regulations are sufficient or satisfactory. Mutual recognition acknowledges that different national standards can be considered interchangeable for domestic application.
- Cooperation : The process through which regulators or agencies from different legal regimes address disparities and establish clear mandates. Cooperation may involve collective regulatory rules or coordination in designing, implementing, and enforcing regulatory measures.

INTEROPERABILITY INSTRUMENTS

Technical interoperability :

- Mutual consensus through the development and adoption of open standards
- Creation of infrastructure for integrations
- Open source code
- Policy intervention to advance interoperability in a number of specific technology contexts
- Common protocols
- Couplings between hardware and software
- Sharing data between services
- “Adversarial interoperability” - engineering interoperability without its maker’s consent or involvement

UNU Macau

EFFECTIVE ETHICAL INTEROPERABILITY INSTRUMENTS

1) **Shared or common terminology**: countries demonstrate similar ethical concerns in AI safety governance. This alignment lays a foundational basis for ethical interoperability.

- South Korea : emphasizing AI safety, accountability, and transparency; a robust privacy regime.
- China: promotes a human-centric ethical framework for AI, integrating values such as fairness, justice, transparency, accountability, and respect for human dignity into AI design and governance.
- UK: five core principles: safety, fairness, accountability, contestability, and adaptability, ensuring AI decision-making aligns with ethical norms.
- Singapore: key values include safety, accountability, transparency, and protection.

EFFECTIVE ETHICAL INTEROPERABILITY INSTRUMENTS

2) **Compatible cross-institutional and international ethical assessment and accountability mechanisms**:

- Domestically, the four governments combine foundational legislation, industry-specific rules, and adaptive ethical guidelines to build comprehensive AI safety governance ecosystems. These layered systems enable AI development under relatively clearly defined ethical boundaries.
- Internationally, countries align domestic initiatives with global frameworks. These global norms identify and integrate diverse national values, providing a common reference for assessing AI practices, facilitating cross-border regulatory coordination, and promoting mutual trust and interoperability in AI among countries.

3) **Multi-stakeholder engagement mechanism:**

Countries have adopted inclusive governance models incorporating diverse stakeholders' voices.

EFFECTIVE REGULATORY INTEROPERABILITY INSTRUMENTS

Standardisation: States are aligning their regulatory frameworks with globally or regionally recognised normative benchmarks (guidelines, rules or practices etc) to ensure consistency and trust across borders. These benchmarks are:

a. Cross-border Data Flows

EU's General Data Protection Regulations (GDPR) and its mechanisms including adequacy decision, "continuity of protection" principles, data classification etc; G7's "Data Free Flow with Trust" (DFFT) framework; OECD's Data Governance Principles; UNESCO Recommendation on the Ethics of AI; The new Global CBPR (Cross-Border Privacy Rules) evolved from the APEC's Cross-Border Privacy Rules system, is developing towards a global data transfer framework and certification processes for personal data transfer. All four jurisdictions are either its members or observers (China).

b. Education

UNESCO's Recommendation on the Ethics of AI and Global Education Coalition on AI, prioritise student well-being, educational equity, and the preservation of teachers' roles as well as principles of fairness and inclusion.

c. Autonomous Vehicles

United Nations Economic Commission for Europe (UNECE) vehicle regulations (UNECE WP.2- The UNECE World Forum for Harmonization of Vehicle Regulations)

EFFECTIVE REGULATORY INTEROPERABILITY INSTRUMENTS

Harmonization: States have established unified regulations through bilateral or multilateral agreements.

a. Cross-border Data Flows

China has applied to join digital trade agreements with Singapore, Chile, and New Zealand. South Korea forges digital trade agreements and partnerships with countries like Singapore, Vietnam, and the UK makes commitments on cross-border data flows, digital trust, and cooperation on AI ethics.

b. Autonomous Vehicles

UK harmonises AV regulations through the UNECE WP.29 (The UNECE World Forum for Harmonization of Vehicle Regulations)

c. Education

The UK is negotiating bilateral and multilateral agreements (such as the CPTPP) that promote EdTech exchange etc.

EFFECTIVE REGULATORY INTEROPERABILITY INSTRUMENTS

Mutual recognition: Jurisdictions are increasingly accepting each other's regulatory mechanisms to reduce compliance burdens.

a. Cross-border Data flows

GDPR-style adequacy mechanisms have been accepted and actively adopted in countries like South Korea, Singapore and the UK.

UK Adequacy Decision recognises the EU/EEA, Japan, South Korea, Canada, and others as adequate partners.

South Korea aims to set up adequacy decisions with the UK, Singapore, and other jurisdictions. South Korea and its partners are exploring mutual recognition of AI audits and certifications.

China and Germany have signed a Memorandum of Understanding (MoU) on China-Germany Cooperation in Cross-Border Data Flow, facilitating cross-border data exchange for enterprises

EFFECTIVE REGULATORY INTEROPERABILITY INSTRUMENTS

Cooperation: Regulators from different states cooperate to overcome the disparity of different regulatory regimes via multilateral dialogue forums, dedicated working groups, joint research and sharing best practices or pilot experience.

a. Cross-border data flows

AI Summit dialogues held in the UK, South Korea and France; China, South Korea and UK participated in multilateral forums including OECD Working Party on Data Governance and Global Partnership on AI (GPAI)'s Data Governance Working Group and Global Privacy Assembly

b. Autonomous Vehicles

China and Germany have signed the Joint Statement of Intent on Cooperation in the Field of Autonomous and Connected Driving, and will jointly develop vehicle-to-everything (V2X) technology; UK participates in Horizon Europe and partnership initiatives like the G7 Transport Ministers' declarations to develop joint research, regulatory alignment and shape global best practices.

c. Education

China participates in UNESCO's Global Education Coalition on AI, and has Sino-foreign university joint research on AI safety in education. These collaborations focus on key issues such as mitigating algorithmic bias and protection of minors' data. UK collaborates in digital and AI skills through the OECD and OECD AI Policy Observatory, and cross-border research in pedagogical innovations and AI progress, and shares global best practices to meet its strict safety and ethical standards

EFFECTIVE TECHNICAL INTEROPERABILITY INSTRUMENTS

Technical standardisation: by adopting common standards across jurisdictions, across software, hardware components, and platforms

a. Cross-border Data Flows

- China adopts global technical standards such as TLS 1.3 for encrypted transmission and ISO/IEC 27701 for privacy management, its GB/T 43697-2024 standard for Data Classification and Grading Rules aligns with international data risk assessment methodologies.
- South Korea adopts international data format standards such as IEEE Learning Data standards for ed-tech, or ADAS/AD sensor data formats for vehicles so that when data crosses borders it remains interpretable and usable by foreign AI systems without needing error-prone conversion.
- UK complies with frameworks like ISO 27001 for information Security, UNECE regulations for safety, ISO standards' specific measures designed for autonomous functionality and ISO/IEC 27701 for privacy management

EFFECTIVE TECHNICAL INTEROPERABILITY INSTRUMENTS

b. Autonomous Vehicles

China's grading standards GB/T 40429-2021 adopts the UNECE WP.29's automation classification to ensure compatibility with global AV terminology. Functional safety standards GB/T 34590 series are equivalent to ISO 26262, and GB/T 38667-2020 (SOTIF) and GB 44495-2024 (Vehicle Cybersecurity) references ISO/SAE 21434. Together they set the requirements for system design, testing, and validation.

UK follows global standards such as ISO 26262 (electronic systems), ISO 21448 (intended functionality safety), and ISO/SAE 21434 for cybersecurity. The UK enforces UNECE WP.29 regulations including Regulation 157 (Automated Lane Keeping Systems) and Regulation 155 (cybersecurity). BSI's programme promotes adopting emerging standards like ISO 34503 on operational design domains, with PAS 1883:2025 guiding local implementation.

South Korea works with UNECE WP.29 on vehicle regulations.

c. Education

South Korea's reference emerging IEEE/ISO AI-in-education standards. China contributes to ISO/IEC JTC1 SC36 (Learning Technologies) and sharing experiences from the National Smart Education Platform.

EFFECTIVE TECHNICAL INTEROPERABILITY INSTRUMENTS

Collaborations: Joint efforts in research and standard-setting to enhance global interoperability.

a. Cross border data flows

South Korean experts participate in ISO/IEC and ITU working groups on AI and data standards, developing common data schemas, metadata standards, and ontologies. They also contribute to the OECD's work on AI system risk classification and ISO/IEC JTC 1/SC 42 on AI which develops standards for the AI lifecycle. These technical standards facilitate cross-border acceptance of AI products.

b. Autonomous Vehicles

China participates in UNECE WP.29 on global AV regulations and ISO/IEC JTC1 on AI safety standards, with standards such as GB/T 40429-2021 being referenced in international AV standardisation discussions. China collaborates with Germany on joint Vehicle-to-Everything (V2X) technology R&D aligning with global efforts to test cross-infrastructure AV interoperability.

EFFECTIVE TECHNICAL INTEROPERABILITY INSTRUMENTS

Creation of Infrastructure for Integrations

South Korea invests in technical infrastructure for cross-border data flows to secure international network links and cloud arrangements under the principle of “secure data corridors” – a hardened pipeline for data exchange with designated certified cloud centres in Korea and other countries with encrypted VPN connections and mutual audits. This infrastructure, combined with common technical standards and certifications, aims to provide a backbone for trustworthy AI collaboration internationally. South Korea can enforce its rules and lower practical barriers for companies to comply. It also fosters innovation since companies can integrate into global data ecosystems using standardised APIs and certifications rather than negotiating one-off arrangements.

Singapore’s AI Verify Toolkit helps companies assess the responsible implementation of their AI system against internationally recognised AI governance principles. The framework is aligned with other international frameworks such as those from EU, G7, OECD, and the US.

INTEROPERABILITY -INTEROPERABILITY BARRIERS

Ethical interoperability barriers

Challenges in Algorithmic Transparency and Accountability for Safety-critical Systems

Limited Universality of International Ethical Frameworks

Voluntary Nature of Ethical Principles

Uneven Maturity and Implementation of AI Standards

Technical Standard Interoperability Barriers

Overlapping Efforts Among International AI Standard Bodies

Focus on Deployment Safety and Assurance:
Gaps in Catastrophic/Frontier Risk Mandates:

Regulatory interoperability barriers

Lack of Global Regulatory Standardization

Geopolitical Tensions Undermine Frontier AI Safety Collaboration

Regulatory Harmonization through Digital Trade Agreements:

Prioritising Political and Commercial interest of Interoperability Over Legal Harmonisation

Diverse Liability Frameworks for Autonomous Driving

Fragmented Liability and Data Protection in Cross-Border Data Flows

Limited International Alignment in AI Safety in Education

Need to Invest in Digital Public Infrastructure

UNU Macau

Policy Recommendations-Ethical Interoperability

Leveraging existing effective interoperability instruments at both global and national levels

Promote Ethical Self-Certification Reports

the UN takes the lead in convening, coordinating, and overseeing the development of the AI Ethical Self-Certification Report mechanism, while national governments, regional bodies, or organisations would be responsible for submitting their respective self-certification reports.

Uphold the UN System as the Primary Forum for AI Ethics Deliberation

the UN to propose more concrete measures for global AI ethical governance, for example, introducing AI ethics indicators or establishing a global AI ethical monitoring dashboard.

Advance the Global AI Ethics Framework

Current efforts often rely on broad principles, non-binding guidelines, and voluntary commitments. We urge platforms such as the Global Dialogue on AI Governance (GDAIG) and the Independent International Scientific Panel on AI (SPAI) to lead in propose universally applicable ethical standards that can serve as the basis for concrete, operational, and potentially binding international policy instruments, supported by a set of clear, small-scale Key Performance Indicators (KPIs) to track progress, identify gaps, and enable regular cross-country and cross-organizational AI safety review.

Policy Recommendations -Technical Interoperability

Promoting Interoperability by Design: Countries are encouraged to align their national standards with established and emerging international benchmarks. Regulators can support this effort by developing sector-specific checklists or AI audit requirements aligned with international frameworks or relevant UN guidance. Such alignment enables AI industries to build systems that are compliant by design, facilitating smoother integration into global marketplaces and supply chains. It also helps ensure that AI technologies are developed and deployed safely, ethically, and interoperable across borders, supporting a more cohesive and trustworthy international AI ecosystem.

Call for International Consensus-Driven AI Standards and Avoiding Duplication: To ensure safe, ethical, and interoperable AI development, countries should support the creation of international consensus-driven standards and avoid unnecessary duplication of standardization efforts.

Prioritizing the Development of Dedicated AI-in-Education Standards: This could be achieved by launching *Technical Specifications for AI Educational Tools*, covering key areas such as content quality, algorithmic fairness, and data security.

Policy Recommendations -Technical Interoperability

Prioritizing Standards Interoperability at the Security Layer: Interoperability efforts should focus more on the security layer than the technical layer. The security layer protects AI systems and data from harm - including accidents, misuse, and cyber threats - and requires common safety and security standards that can be applied across sectors and borders.

Call for Scenario Planning for AI-Related Catastrophic Risks and Improved Regulatory Forecasting: addition to focusing on AI deployment safety and assurance through tools like the AI Verify Toolkit, AI Assurance Sandbox, and mechanisms for testing and certification, governments and regulators must expand their attention to long-term and high-impact risks. scenario planning and regulatory forecasting for model-level controls are essential. This includes oversight of foundational training data, intrinsic safety features of general-purpose AI, and mitigation strategies for catastrophic risks such as loss of control or dual-use threats (e.g., bioweapons or advanced cyber offense). Most critically, soft ethical guidance for high-risk AI must be transformed into enforceable and auditable obligations. These should cover: Data governance; model evaluation; pre-deployment and post-deployment testing; incident reporting and corrective actions.

Policy Recommendations -Regulatory Interoperability

➢ GOVERNANCE ARCHITECTURE AND COORDINATION

These recommendations focus on building institutional structures and international cooperation mechanisms to support regulatory interoperability.

- **Establish a Multilateral System for Coordinated AI and Data Governance:** States negotiate the creation of a multilateral system with a designated institution or mechanism to coordinate a common approach. This system should uphold the United Nations (UN) Charter's principle of sovereign equality among all Member States, as applied to national actions in cyberspace.
- **Establish a Coherent National Entity for Global Engagement :** To effectively engage with UN bodies and global partners, countries should formalise a coherent national entity to design a unified national approach. a *National AI Safety Coordination Council* to address regulatory inconsistencies and represent national interests in global AI governance forums.
- **Support Inclusive Multi-Stakeholder Engagement for AI Safety Governance:** Effective AI safety governance depends on inclusive mechanisms; establish advisory bodies with the authority to provide policy input and oversee the safe use of AI technologies and data. Supporting independent AI safety institutes to offer expert, unbiased advice to guide decision-making. Independent International Scientific Panel on AI (SPAI) and the Global Dialogue on AI Governance (GDAIG) provide scalable models for inclusive governance. Both bottom-up participation and top-down coordination in shaping AI safety governance.
- **Enhance Public Engagement to Strengthen International AI Governance :** The long-term success of AI governance depends on public trust and the democratic values emphasized in UN guidelines. To cultivate this trust, countries should launch national dialogues on AI's societal impacts, engaging educators, industry leaders, technical experts, and civil society early through schools, public forums, and dedicated task forces.

Policy Recommendations -Regulatory Interoperability

➢ STANDARDS, BENCHMARKS, AND TRANSPARENCY

These aim to improve interoperability through shared benchmark, transparency, and accountability mechanisms.

- **Launch a Global Benchmark for AI Safety and Security:** To support regulatory alignment among states, there is an urgent need to establish a global benchmark for AI safety and security to implement the Global Digital Compact (GDC) and the two recent UN resolutions on AI governance.
- **Enhancing Transparency and Accountability in AI Safety Governance:** To strengthen public trust and improve oversight, countries are encouraged to develop clear AI safety key performance indicators - such as autonomous vehicle accident rates, data breach incidents, and public trust levels. Publishing an annual national *AI Safety Report* can help track progress, inform the public, and guide the refinement of policies and governance frameworks. Governments should also enhance transparency in policymaking by publicly sharing the reasoning, standards, and evidence behind AI-related regulations. Regular AI policy evaluations, including annual white papers or audits of key programs, should be conducted and made publicly available to ensure accountability and continuous improvement.

Policy Recommendations -Regulatory Interoperability

> TECHNICAL AND LEGAL INFRASTRUCTURE FOR INTEROPERABILITY

These address the development of interoperable systems, legal harmonization, and data governance mechanisms.

- **Promote Adaptation and Expansion of Data Interoperability Mechanisms**: In the absence of a unified global framework for cross-border data transfers, it is essential to encourage the adoption of effective mechanisms for standardization, harmonization, mutual recognition, and international cooperation. Potential instruments include: Alignment with global or regional benchmarks; bilateral and multilateral agreements (e.g., digital trade agreements, UNECE WP.29); GDPR-style adequacy mechanisms; mutual recognition of AI audits, data protection certifications, and test results; multilateral dialogue forums and dedicated working groups, joint research initiatives and sharing of best practices or pilot experiences. These mechanisms can foster interoperability, reduce redundant compliance burdens for enterprises, mitigate risks, and lower cross-border frictions. They also enhance portability of compliance, supporting data free flow with trust.

Policy Recommendations -Regulatory Interoperability

> TECHNICAL AND LEGAL INFRASTRUCTURE FOR INTEROPERABILITY

These address the development of interoperable systems, legal harmonization, and data governance mechanisms.

- **Additional Protections to Mitigate Data Flow Risks:** In addition to voluntary international mechanisms such as the Global CBPR System and ASEAN Model Contractual Clauses (MCCs), more robust protections are urgently needed to safeguard cross-border data flows. These enhanced safeguards include: Adoption of ethical principles such as “fundamental rights travel with the data”, “universal privacy and dignity”, and the GDPR’s “continuity of protection” principle; implementation of thorough due diligence as a key risk mitigation measure, ensuring that domestic companies remain accountable for data breaches occurring overseas; and introduction of cross-border data liability insurance to help companies manage financial risks associated with international data transfers. promoting open data flows with trust, embedding security by design, and integrating safety, privacy, and human rights into digital cooperation.
- **Development of Interoperable Digital Public Infrastructure:** States can collaborate and invest in digital infrastructure to enhance global interoperability - particularly through initiatives like standardized digital mapping of road regulations for autonomous vehicles (AVs). Such efforts are essential for harmonizing technological deployment across borders and reducing regulatory fragmentation. They also support Sustainable Development Goal 9 (Industry, Innovation, and Infrastructure) by promoting inclusive infrastructure for technology deployment, fostering innovation, and enabling a more competitive and cooperative global market.

Policy Recommendations -Regulatory Interoperability

> TECHNICAL AND LEGAL INFRASTRUCTURE FOR INTEROPERABILITY

These address the development of interoperable systems, legal harmonization, and data governance mechanisms.

Developing Greater Standardization and Harmonization of Liability Models for Autonomous Vehicles: While a globally uniform liability model for autonomous vehicles may not be foreseeable in the near future, there is a growing recognition of the need for greater international standardization and harmonization. Efforts are underway through international bodies such as the ISO 39003:2023 standard - *Road Traffic Safety (RTS): Guidance on Ethical Considerations Relating to Safety for Autonomous Vehicles* - and the United Nations Economic Commission for Europe (UNECE) WP.29 liability framework. UNECE WP.29's technical regulations and guidance documents help member countries address liability in the event of an accident within their own legal systems. However, no internationally binding legislation currently exists. Given that traffic regulation is embedded in system design, industry stakeholders have called for regulatory harmonization at the international level. Legislation needs to evolve alongside technological advancements, without hindering progress at national borders. In particular, statutory liability rules for fully autonomous (Level 5) vehicle systems must be collectively researched and addressed by states to ensure legal clarity, safety, and accountability.

Policy Recommendations -Regulatory Interoperability

> RESEARCH AND CAPACITY BUILDING

These support long-term interoperability through education, research, and risk management.

Investment in AI Safety, Frontier Risk Management, and Alignment Research Collaborations: Significantly increasing investment and international collaboration in AI safety research is essential for managing the risks associated with advanced AI systems.

Promoting AI Safety in Education: To ensure the safe and responsible use of AI, the general public and workforce need a stronger understanding of its capabilities, risks, and ethical implications. This calls for a holistic approach to AI safety in education. AI literacy should be integrated into school curricula - not only within computer science classes but also through broader digital citizenship education. Teacher training must be expanded to include AI safety and ethics, supported by potential accreditation schemes. Evaluation mechanisms are needed to monitor bias and efficacy in AI-related educational content. Education systems should also broaden reskilling programmes for workers likely to be impacted by AI-driven job displacement, helping to prevent social harm. Safety and ethics training should be embedded in both initial teacher preparation and ongoing professional development. UNESCO Recommendations on AI in Education and the IEEE 3527.1™ Standard for Digital Intelligence (DQ) offer valuable guidance for countries seeking to align their educational strategies with International benchmarks.

Conclusion

- The future of AI safety governance is evolving towards an evidence-based, outcomes-oriented model that complements principle-led frameworks. This shift reflects a growing international consensus and increasing alignment with emerging global standards.
- To sustain momentum, policymakers must prioritize deepening normative specificity, expanding the adoption of interoperable standards, strengthening data governance mechanisms, and collaboratively investing in and conducting research into technical, institutional, and AI literacy capacity building.
- Achieving AI safety and interoperability is a dynamic and iterative process. In this process, public trust defines the foundation, interoperable ethics guide strategic direction, regulations translate values into enforceable rules, standards and enforcement drive implementation, and international cooperation amplifies impact.
- Together, these elements form a resilient, inclusive, and globally aligned AI safety governance ecosystem.

UNU Macau

Acknowledge

The publication has greatly benefited from the research assistance of Songruowen Ma and Eduarda Mello

We appreciate the comments and feedback from Antonella Maia Perini, Gurjit S. Sandhu, Christopher David Jones, Alan Duncan King, Raymond Forbes and Rolf H. Weber

This work is made possible through the generous funding of SenseTime

UNU Macau