Background Paper

Human Rights & Artificial Intelligence

Virginia Dignum[1], Rachele Carli[1], and Tang Yingxia[2]

[1] AI Policy Lab, Umeå University

[2] Human Rights Research Centre, Nankai University

The 23rd Informal ASEM Seminar on Human Rights 29-31 October 2025

CO-ORGANISED BY

HOST

MINISTRY OF FOREIGN AFFAIRS
OF DENMARK

SUPPORTED BY

Contents

Executive Summary

This background paper has been prepared to inform the 23rd Informal ASEM Seminar on Human Rights on the theme of Artificial Intelligence (AI) and Human Rights. It provides an overview of the main opportunities and risks posed by AI across Asia and Europe, structured around three thematic foci: privacy and data protection, equality and non-discrimination, and remedies and access to justice. It also identifies future directions and concrete opportunities for Asia–Europe cooperation in the governance of AI.

AI technologies are rapidly transforming social, economic, and political life. They offer significant benefits in areas such as healthcare, education, and public administration, but also create acute risks for fundamental rights. The use of AI in surveillance, welfare allocation, recruitment, and online platforms has raised pressing concerns over mass data collection, profiling, systemic bias, and limited access to redress. These challenges are amplified by the unevenness of regulatory regimes across ASEM countries and by the transnational nature of AI-related harms.

The paper identifies three core sets of challenges, to be the focus of the discussions during the ASEM seminar:

- **Privacy and data protection:** AI systems rely on the large-scale collection and processing of personal data, creating risks of mass surveillance, opaque profiling, and weak safeguards. While Europe has consolidated protections through the GDPR, Convention 108, and the AI Act, Asian states show more diverse approaches, ranging from binding frameworks in India, South Korea, and Vietnam, to soft-law initiatives such as ASEAN's Guide on AI Governance and Ethics and Singapore's AI Verify toolkit.

- **Equality and non-discrimination:** AI often reproduces or amplifies social biases, with significant consequences in welfare systems, employment, credit scoring, and surveillance. Europe frames bias as a rights violation prohibited under binding instruments, while Asian states have adopted a patchwork of sectoral regulations and judicial interventions. Systemic remedies remain limited across the region.

- **Remedies and access to justice:** Effective redress mechanisms for AI-related harms remain fragmented. Europe provides stronger procedural and institutional safeguards through DPAs, ombuds institutions, and courts, though enforcement gaps persist. In Asia, remedies are uneven, with early experiments such as disclosure duties (China), complaint mechanisms (Philippines), and voluntary Human Rights Impact Assessments (South Korea).

The comparative analysis highlights key divergences: Europe tends toward comprehensive, enforceable frameworks, while Asia shows heterogeneous and fragmented approaches, often balancing human rights with state control and economic development. Despite these differences, the underlying risks are shared, and governance challenges are convergent.

Looking ahead, the paper emphasises that effective AI governance requires moving beyond high-level ethical principles towards enforceable safeguards, algorithmic

accountability, and inclusive participation. It underlines that regulation and innovation are not in conflict: robust governance is essential to building trust, legitimacy, and sustainable adoption of AI technologies.

Finally, the paper identifies concrete opportunities for Asia–Europe cooperation within ASEM, including:

- Establishing an ASEM Observatory on AI and Human Rights.
- Launching joint training programmes for regulators, judges, and civil society.
- Piloting cross-border AI audits or certification schemes integrating human rights safeguards.
- Supporting multi-stakeholder dialogue platforms to ensure inclusive participation.

Taken together, these actions provide a roadmap for ASEM partners to translate shared commitments into practice. By embedding human rights at the core of AI governance, Asia and Europe can demonstrate global leadership in aligning technological innovation with the protection of fundamental rights and democratic values.

1 Introduction

1.1 Background and Core Concepts

In this paper, the term Artificial Intelligence (AI) is used as an umbrella term covering a broad family of computational techniques, including machine learning, natural language processing, computer vision, and decision-support systems, rather than a single technology. This reflects the way AI is framed in policy and human rights contexts, where the focus is on governance, safeguards, and impacts on rights, rather than on technical taxonomies.

In all its different meanings and approaches, AI technology is rapidly reshaping societies across Asia and Europe, offering both significant opportunities and complex challenges for the protection and promotion of human rights. As governments, businesses, and civil society actors increasingly rely on AI systems to make or support decisions in areas such as healthcare, education, law enforcement, and public administration, concerns about transparency, accountability, fairness, and non-discrimination have moved to the forefront of human rights discourse.

This background paper, prepared for the 23rd Informal ASEM Seminar on Human Rights, examines the evolving intersection between AI and human rights within the Asia-Europe context. It aims to foster dialogue among policymakers, academics, technologists, and civil society actors from both regions by providing an overview of the emerging regulatory landscape, key thematic issues, and shared challenges and opportunities for cooperation. In particular, it focuses on three thematic pillars central to the Seminar: privacy and data protection, equality and non-discrimination, and access to remedies.

AI systems are not inherently neutral; their design and deployment can either advance or undermine fundamental rights, depending on the values embedded in their governance. On the one hand, AI can help expand access to services, improve public sector efficiency, improve early warning systems for human rights violations, and support the realisation of economic, social, and cultural rights through data-driven insights. On the other hand, serious risks arise from opaque algorithmic decision-making, embedded biases in training data, lack of meaningful oversight, and inadequate legal and institutional safeguards. These risks threaten core rights such as privacy, freedom of expression, equality, and access to justice. They also highlight the importance of addressing intersecting forms of discrimination, such as those based on gender, race, disability, or socio-economic status, which remain central to equality and non-discrimination obligations. A nuanced understanding of this dual potential is essential for developing governance frameworks that harness the benefits of AI while preventing and mitigating its harms.

Similarly, it could be highlighted that the intersection of human rights with the challenges of technological development is twofold: fundamental rights serve both as a normative framework to guide the development of AI systems and as a body of law whose full realisation may be actively fostered and supported through AI.

The international human rights regime provides a universally recognised set of principles and obligations, including rights to privacy, non-discrimination, freedom of expression, and access to remedy, that serve as a foundation for responsible tech-

nological innovation. Viewing human rights as a framework for AI development entails embedding these standards at every stage of the technology's lifecycle, from design to deployment and ex post supervision. This means that AI systems should be conceived and implemented with the explicit purpose of respecting, protecting, and promoting human rights [154]. Such a rights-based approach requires systematic assessment, transparency, accountability mechanisms, and the inclusion of stakeholders — especially impacted communities and minorities — in governance and decision-making processes. Human rights impact assessments are meant exactly for this: they enable developers and policymakers to anticipate, identify, and mitigate potential harms before systems are deployed. In doing so, they shift the focus from mere legal compliance to the proactive advancement of human dignity, fairness, and social equity. Table 1 provides an overview of existing Human Rights Impact Assessment frameworks.

Conversely, AI technologies possess significant potential to support and strengthen the body of human rights law itself. In fact, despite their universality, the problem of effective access to instruments for the protection of fundamental rights, or even their full enjoyment, is far from being resolved at the international level. Machine learning, natural language processing, and large-scale data analytics can enhance legal discovery, documentation of abuses, and access to justice. AI-powered platforms can be utilized to monitor rights violations, increase legal literacy, and facilitate reporting and redress, thereby contributing to the realization of fundamental rights in practice. Furthermore, AI-powered remote learning or telemedicine tools can increase access to education or healthcare for all those who find themselves in circumstances — whether personal or determined by external factors — that make it difficult to access or fully enjoy these essential rights. In order for this impact to be effective and truly beneficial, it is critical that such systems are developed and deployed in light of the principles highlighted in the previous paragraph.

Therefore, it follows that a robust human rights perspective on AI both directs the ethical and legal boundaries of system design and positions AI as a technological ally in the global effort to advance and protect equity.

To better understand and address these elements, the paper begins by introducing foundational concepts relevant to AI and human rights alike. Definitions and guiding principles are drawn from international and regional legal frameworks, including the Universal Declaration of Human Rights (UDHR), the UN Guiding Principles on Business and Human Rights (UNGPs), the Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, and the European Union's Artificial Intelligence Act. These documents provide important normative benchmarks and legal tools for assessing the human rights implications of AI technologies.

Following this conceptual foundation, the paper is structured into four main sections: (1) a review of international and regional frameworks for human rights and AI governance; (2) thematic analyses focusing on privacy and data protection, equality and non-discrimination, and access to remedies; (3) forward-looking considerations on integrating human rights into AI governance; and (4) a discussion of future trends and potential areas for Asia-Europe collaboration.

By situating the debate within the ASEM framework, this paper highlights the importance of cross-regional dialogue and shared responsibility. It calls for a rights-based approach to AI governance that is inclusive, participatory, and context-sensitive,

Table 1: Examples of human rights impact assessment (HRIA) frameworks relevant to technology and AI.

Danish Institute for Human Rights (DIHR) – HRIA Guidance and Toolbox
Comprehensive toolkit with step-by-step guidance for planning, conducting, and reporting human rights impact assessments.
https://www.humanrights.dk/tools/human-rights-impact-assessment-guidance-toolbox
UN Guiding Principles on Business and Human Rights – Due Diligence (OECD)
Global standard requiring businesses to identify, prevent, mitigate, and account for human rights impacts through systematic due diligence.
https://www.oecd.org/en/topics/due-diligence-for-responsible-business-conduct.html
Global Network Initiative (GNI) – Assessment Toolkit
Guidance for ICT companies to assess risks to privacy and freedom of expression, with independent accountability mechanisms.
https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf
Institute for Human Rights and Business (IHRB) – ICT Sector Guide
Sector-specific guidance for ICT companies on operationalising the UNGPs, with practical advice and risk mapping.
https://op.europa.eu/en/publication-detail/-/publication/ab151420-d60a-40a7-b264-adce304e138b
OECD – Due Diligence Guidance for Responsible Business Conduct
Risk-based due diligence framework covering human rights, environment, labor, and governance, widely applied in digital and AI contexts.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2018/02/oecd-due-diligence-guidance-for-responsible-business-conduct_c669bd57/15f5f4b3-en.pdf
European Union – Fundamental Rights Impact Assessment (FRIA)
AI Act Art. 27 requires deployers of high-risk AI systems to assess and mitigate risks to fundamental rights prior to deployment.
https://artificialintelligenceact.eu/article/27/

grounded in legal standards and informed by ethical considerations. Beyond dialogue, the paper foreshadows concrete opportunities for Asia–Europe collaboration—such as joint capacity-building, harmonised audit practices, and shared institutional mechanisms—to advance a coherent and rights-based approach to AI governance. Ultimately, the paper seeks to support ASEM partners in developing coherent, just, and future-oriented responses to the evolving challenges posed by AI.

2 International and regional protection

International and regional protection of human rights in the context of AI has become an urgent political issue, as the spread of this technology introduces new dimensions of risk alongside opportunities for social progress. The implementation of AI systems, particularly by states and multinational corporations, is now recognised as potentially transformative for societies, but also as a source of threats inherent to fundamental rights enshrined in international human rights law. Given the nature of AI systems and their development and training techniques, it is now well established that the implications of their use often transcend national borders. Furthermore, it would not be possible to protect fundamental rights without due reference to the international disciplines that address and protect them even outside the mere digital context.

In this regard, the Universal Declaration of Human Rights (UDHR) represents the foundational instrument for the protection of fundamental rights, proclaiming a common standard for the recognition and safeguarding of these rights for all individuals and nations. The Declaration, adopted by the United Nations General Assembly in 1948 in the aftermath of World War II, aimed not only to enshrine in a legal document the rights that should be recognised for all human beings by virtue of their humanity, but also to establish the right to effective remedies in the event of violations. In its formulation, the UDHR has served as both a starting point and a source of inspiration for many national constitutions that emerged thereafter. Among the rights most prominently featured — and frequently invoked in international legislation, including in the context of AI governance — are the right to human dignity, as the cornerstone of the human rights framework, as well as the rights to integrity, equality, and the free expression of one's self, thoughts, beliefs, and identity. Alongside the UDHR, it is essential to highlight both the International Covenant on Civil and Political Rights (ICCPR) and the Convention on the Rights of the Child (CRC) as benchmarks for the protection of such rights.

The ICCPR, in force since 1976, legally binds its signatory states to guarantee the full range of civil and political rights to all individuals. While many of these rights reflect those recognized in the UDHR, their explicit specification and separate articulation were intended to reinforce respect for dignity and equality, particularly in the face of political or administrative dissent within individual states. In this respect, the Preamble's reference to "freedom from fear" is especially noteworthy.

The CRC, adopted in 1989, explicitly addresses the protection and promotion of fundamental rights for individuals under the age of 18. It marked a significant milestone — not least because of its legal force — in the reconsideration of the child as an individual with full rights, whose dignity, identity, and entitlements must be protected independently of those who act as their guardians. Supporting this view, the CRC enshrines the principle of the best interests of the child (Article 3) and the right of the child to be heard in all matters affecting them (Article 12).

This body of norms serves as the starting point for discussions on the inalienable rights at stake in the regulation and development of emerging technologies, including evaluations of which rights AI systems might primarily help to promote. Building upon these principles and rights, numerous legal and governance documents — at both inter-

national and regional levels — have since been developed to more specifically address the interaction between human rights and AI systems.

The following sections present some of the most significant strategies, initiatives, and legal instruments relevant to this intersection. Extended details, including full texts and comprehensive references, are provided in Appendix A.

2.1 Human Rights and AI at the International Level

At the international level, several instruments and initiatives provide normative guidance on the relationship between AI and human rights. These include both binding human rights treaties and non-binding standards that are increasingly shaping global practice. However, a key gap remains the absence of a comprehensive, universally binding legal instrument specifically dedicated to AI and human rights.

2.1.1 United Nations Special Rapporteurs

The system of UN Special Rapporteurs provides independent, expert analysis on human rights. While there is not yet a dedicated mandate for AI, several Rapporteurs have already raised concerns about its use.

The Special Rapporteur on Human Rights while Countering Terrorism highlighted the use of AI for mass surveillance, particularly targeting journalists and activists, and called for a moratorium until safeguards for privacy and freedom of expression are in place [7]. The Special Rapporteur on the Right to Privacy issued report A/78/310 stressing the principles of transparency and explainability in AI data processing [118], and later reports on neurodata (2025) reiterated these concerns. Similarly, the Special Rapporteur on the Right to Education underscored both opportunities (inclusion, disability support) and risks (educational disparities, alienation of teachers) linked to AI in education [151].

These interventions illustrate the increasing attention given by existing mandates to AI-related risks. They also support calls for a dedicated Special Rapporteur on AI and Human Rights, to provide consistent leadership, interdisciplinary expertise, and consolidated guidance at the UN level.

2.1.2 International Covenant on Economic, Social and Cultural Rights

The ICESCR (1966) is one of the three foundational treaties of the International Bill of Rights. Article 15(b) establishes the right to "enjoy the benefits of scientific progress and its applications" [84]. This provision can be interpreted as guaranteeing equitable access to AI technologies while also obliging states to mitigate harms through transparency, oversight, and remedies [139].

In the context of AI, this means that states must ensure that new technologies support, rather than undermine, the enjoyment of economic, social, and cultural rights. Examples include access to education, health care, and social protection through AI-enabled services. At the same time, states must establish mechanisms for accountability, ensuring individuals can contest algorithmic decisions that affect their rights.

2.1.3 Office of the High Commissioner for Human Rights

The OHCHR has been central in identifying AI-related human rights challenges. Reports and Human Rights Council sessions have highlighted risks arising from biometric surveillance (e.g., facial recognition), predictive policing, and algorithmic discrimination [2, 31]. The opacity of such systems (the "black box" problem) makes accountability and access to remedies particularly difficult.

The OHCHR has also stressed the risks of AI-driven content moderation, which may suppress freedom of expression online. To address these challenges, OHCHR has reaffirmed three pillars: (1) human rights as the normative framework for AI development, (2) the need for international cooperation, particularly involving the Global South, and (3) the importance of timely intervention before harmful technologies become widespread.

Recent initiatives include updated interpretative guidance on the UN Guiding Principles on Business and Human Rights (UNGPs), emphasising due diligence throughout the AI lifecycle and stronger stakeholder engagement [122]. In 2025, OHCHR also prioritised the Global Digital Compact, which proposes establishing an Independent International Scientific Panel on AI and a Global Dialogue on AI Governance [62, 88].

2.1.4 Organisation for Economic Co-operation and Development (OECD) AI Principles

In 2019, the Organisation for Economic Co-operation and Development (OECD) adopted the first intergovernmental standard for trustworthy AI: the OECD Principles on Artificial Intelligence. These five core principles emphasise inclusive growth, human-centred values, transparency, robustness, and accountability, supported by five policy recommendations on investment, enabling ecosystems, governance, skills, and international cooperation.

By 2024, these Principles were updated to respond to rapid advances in general-purpose and generative AI. Key revisions included explicit reference to environmental sustainability, a stronger focus on systemic risk management, and the reframing of transparency as *contestability* (the ability to challenge algorithmic decisions). Accountability was broadened to cover bias, intellectual property, and labour rights [119].

With more than 47 adherent countries across Europe and Asia, the OECD Principles have become a global reference point for AI governance, though their voluntary nature leaves enforcement to national implementation.

2.1.5 G20 AI Guidelines

At the 2019 Osaka Summit, the G20 endorsed AI principles inspired by the OECD, emphasising fairness, transparency, accountability, privacy, and the rule of law [69]. Unlike the OECD Principles, however, the G20 Guidelines function primarily as a political declaration rather than a comprehensive governance framework.

They carry diplomatic weight by aligning major economies around high-level commitments, but lack specific implementation or monitoring mechanisms [143]. Their scope is also narrower, focusing on broad values rather than detailed standards, and omitting issues such as environmental sustainability or sector-specific safeguards.

2.1.6 Global Partnership on AI

The Global Partnership on AI (GPAI), launched in 2020 and hosted by the OECD, represents a multi-stakeholder effort to operationalise responsible AI. With participation from over 20 countries, alongside civil society, academia, and industry, GPAI works through expert working groups on responsible AI, data governance, future of work, and innovation.

GPAI draws heavily on the OECD AI Principles and the UNGPs, seeking to bridge theory and practice by embedding human rights in AI governance. It promotes inclusive stakeholder participation and accountability across the AI lifecycle. However, as a voluntary initiative, GPAI's outputs take the form of reports and recommendations rather than binding rules.

2.1.7 G7 AI Principles and Code of Conduct

In October 2023, the G7 launched the Hiroshima Process, adopting eleven voluntary principles and a Code of Conduct for advanced AI systems. These principles stress risk-based management across the AI lifecycle, including pre-deployment impact assessments, post-deployment monitoring, transparency, security, and incident reporting. They also highlight content authentication, research prioritisation, and support for international technical standards.

The Hiroshima Process represents an important step toward harmonising AI governance among major economies, embedding risk management and human rights considerations. However, the voluntary and transitional nature of the principles limits their enforceability. Their main value lies in shaping national approaches and in building alignment with OECD and GPAI processes.

2.1.8 UNESCO Recommendation on the Ethics of AI

The UNESCO Recommendation on the Ethics of AI, adopted in 2021 by 194 Member States, is the first global normative instrument dedicated specifically to AI [161]. It establishes human dignity as the central guiding principle and calls for inclusivity, gender equality, environmental sustainability, and education for responsible AI [116, 110, 165].

The Recommendation urges states to conduct ethical impact assessments of high-risk AI systems, develop national capacities, and ensure public participation in AI governance. Despite its ambition, the Recommendation remains voluntary and largely preventive, focusing on early stages of AI development. It does not fully address adaptive or generative AI systems that evolve after deployment.

2.1.9 IEEE Ethically Aligned Design

The Institute of Electrical and Electronics Engineers (IEEE) has developed a comprehensive framework as a result of the Global Initiative on Ethics of Autonomous and Intelligent Systems: the Ethically Aligned Design (EAD) principles. The main objective is to provide guidance for the development and deployment of autonomous systems and AI systems that promote human rights.

In particular, EAD makes it explicit that AI must be implemented and operated to promote and protect rights such as the right to life, safety, privacy, equality, and freedom of expression. Special efforts must be made to limit the risk of discrimination based on race, gender, religion, disability, sexual orientation, and other individual characteristics as a result of algorithmic operations. To this end, care must be taken to foster human oversight and human agency, putting in place all necessary measures to prevent manipulation and coercion through AI systems. The IEEE also encourages the drafting of governance frameworks that promote the building of public trust in technology, including by ensuring that AI system outcomes can always be traced back to their source of accountability.

This initiative has been influential in guiding industry and policies towards human rights-based approaches in AI system design, with particular emphasis on de-biasing and transparency. Nevertheless, the effective implementation of these guidelines encounters some difficulties, primarily related to technical complexity. In fact, embedding ethical principles from design to actual deployment of AI requires ethical risk modelling, algorithmic audits, adversarial testing, and other methodologies to ensure transparency, which are difficult to operationalise. Moreover, the claim for constantly verifiable accountability — while certainly commendable and worthy of further study — may be difficult to implement in practice, due to the still high level of unpredictability in AI behaviour in real-life scenarios. While establishing an interdisciplinary ethics review board could be a valid option, organisations may find it difficult to implement consistently due to the related need to ensure diverse stakeholder involvement, constant and continuous commitment, and clear accountability measures, which are difficult to set based on EADs alone. Furthermore, as these are voluntary guidelines, they lack enforcement structures, which could lead to uneven adoption and pressures due to the need for companies to meet efficiency and financial expectations in order to be competitive in the market.

2.1.10 Raoul Wallenberg Institute of Human Rights and Humanitarian Law

The Raoul Wallenberg Institute of Human Rights and Humanitarian Law (RWI) is an independent research and education institute that has increasingly engaged with the intersection of AI and human rights. Its work examines both the opportunities and risks of AI in areas such as healthcare, justice, and social welfare.

RWI stresses the importance of inclusive multi-stakeholder engagement, transparency, and human dignity in the design and governance of AI systems. It also calls for more robust ethical standards and stronger safeguards against algorithmic bias and accountability gaps.

Although not a norm-setting body like the OECD or UNESCO, RWI contributes through research, policy advice, and convening platforms, helping bridge the gap between academic expertise, civil society perspectives, and policy development.

2.1.11 Overview

Taken together, these instruments constitute a growing body of international initiatives. They provide important reference points for rights-based AI governance, but remain

fragmented and uneven in enforceability.

Summary of International Approaches to Human Rights and AI

Instrument / Actor	**Focus / Contribution**	**Limitations**
UN Special Rapporteurs	Highlight AI risks (surveillance, privacy, education); call for dedicated AI Rapporteur	Fragmented, issue-specific mandates
ICESCR (1966)	Right to benefit from scientific progress (Art. 15b); obligations of due diligence and remedies	Broad; requires interpretation for AI context
OHCHR	Reports on surveillance, discrimination, opacity; guidance on UNGPs; Global Digital Compact	Recommendations non-binding; implementation gaps
OECD AI Principles (2019/24)	First intergovernmental AI principles; inclusivity, fairness, transparency, accountability, sustainability	Voluntary; interpretive ambiguities
G20 AI Guidelines (2019)	Political alignment of major economies; based on OECD principles	No monitoring or enforcement mechanisms
Global Partnership on AI (2020)	Multi-stakeholder, practice-oriented; focuses on rights, governance, future of work	Advisory role; limited authority
G7 Hiroshima Process (2023)	Eleven voluntary principles on risk management, transparency, content authentication	Transitional, non-binding
UNESCO Recommendation (2021)	First global normative framework; dignity, inclusivity, sustainability	Preventive, voluntary, limited adaptability
IEEE Ethically Aligned Design (EAD)	Voluntary framework for embedding human rights in AI; stresses oversight, accountability, and non-discrimination	Hard to implement; no enforcement
Raoul Wallenberg Institute (RWI)	Research and policy advice on AI and human rights; emphasises dignity, inclusivity, transparency	Advisory role; non-binding influence

The rapid advancements in generative AI and the emergence of novel applications have introduces a series of potentially unforeseen ethical dilemmas, such as, by way of example only, the dissemination of misinformation, the development of deepfakes, or the misuse of technology, that existing guidelines may not fully address yet, necessitating ongoing updates and research. Furthermore, the effective implementation of AI ethics is contingent upon the education of developers, policymakers, and users regarding the capabilities, risks, and responsible use of AI, a gap that hampers ethical compliance and broader social trust. Concurrently, a global regulatory approach that is equally enforceable and uniformly valid across the world is very complex even just to theorise conceptually, due not only to the variety of legal systems that exist internationally, but also to the cultural and social contexts, and the related historical and philosophical roots that guide the different states and their respective approaches to policy in the various sectors of interest.

However, analysis conducted in this section is useful for highlighting the need for an even stronger international cooperation and potentially a binding global treaty to ensure consistent and coordinated human rights protection in the AI era. This has already begun to take shape through the emergence of emerging global dialogues, like the 2023 AI Safety Summit, the 2024 AI Seoul Summit, and the 2025 AI Action Summit, which have spurred initiatives such as a network of AI Safety Institutes and multi-stakeholder foundations, aimed at democratizing access, setting safety infrastructures, and advancing public-interest AI.

In order to underline the ways in which these international inputs and key principles have been received by ASEM countries, the following sections will examine some of the main AI governance and regulation measures in light of the human rights framework which have been developed in Asia and Europe.

2.2 Human Rights and AI at the regional level: Asia

Asia has emerged as a critical region in the global development and governance of AI, characterized by rapid technological adoption, diverse regulatory approaches, and significant variations in the integration of human rights safeguards. As AI technologies become increasingly embedded in economic and social systems, countries across the region are grappling with how to balance innovation with the protection of fundamental rights. Against this backdrop, South Korea, China, and Japan stand out as particularly instructive case studies, due to their leading roles in AI development, distinct governance models, and active engagement in shaping regional and global AI norms. These three nations represent a spectrum of regulatory philosophies — from South Korea's comprehensive and human-rights oriented framework, to China's approach of balancing development and governance with a people-centred focus, to Japan's principles-based and inclusive governance model. By examining their legislative advances and policy challenges, this subsection aims to illuminate key trends, gaps, and tensions in the evolving relationship between AI and human rights in Asia.

2.2.1 AI Basic Act

South Korea's AI Basic Act, passed on December 26th 2024 and promulgated on January 21st 2025, marks a significant step in the country's efforts to establish a comprehensive regulatory framework for AI . This legislation aims to promote AI technology development while safeguarding human rights and ensuring social responsibility. It constitutes the first comprehensive framework on AI in the Asian region and the second on a global level, following the adoption of the EU AI Act in August 2024 [158].

The core of the AI Basic Act is the establishment of a centralized governance structure, with the National AI Committee leading the policy development. This committee is chaired by the President and composed of government officials and industry experts. This centralized decision-making model ensures consistency and efficiency in policy execution, allowing for coordinated efforts in promoting AI-related initiatives. However, it also raises concerns about governance centralization, which might limit the inclusion of diverse voices from industry stakeholders and reduce the capacity for grass-roots innovation. In a rapidly evolving technological field, a more centralized governance model may impede the flexibility required to adapt to emerging challenges. The AI Basic Act also mandates the creation of a Basic AI Plan every three years, which is led by the Ministry of Science and ICT (MSIT) and aims to support AI research and development. In addition, it provides the legal basis for establishing specialized institutions like the AI Policy Center and the AI Safety Research Institute to promote and enforce AI policies. While this systematized framework helps ensure long-term stability in AI policy, the lack of flexibility in addressing the fast-changing technological landscape could result in delays in policy adaptation, leading to a gap between emerging technological needs and regulatory responses.

One of the key highlights of this Act is its emphasis on human rights protection in AI development. The Act mandates that businesses assess the impact of their AI systems on fundamental rights when deploying high-impact AI technologies. For instance, high-impact AI systems, which may severely affect human life, physical safety, or fundamental rights, are required to undergo additional risk assessments. This provision underscores South Korea's commitment to ensuring that AI technologies do not infringe upon individual freedoms or perpetuate discrimination. However, while the AI Basic Act places significant importance on human rights protection, it does not provide sufficiently clear guidelines on ensuring algorithmic transparency, which is a critical issue in many AI applications. For instance, while businesses are required to inform users when AI is used in their products or services, the Act does not specifically mandate algorithmic transparency or the explainability of AI decision-making processes. In high-stakes areas like facial recognition or social scoring, the lack of transparency could undermine public trust and lead to potential human rights violations, especially if AI decisions are not easily understood or contested by affected individuals [124].

Another distinctive feature of the AI Basic Act is its reliance on voluntary compliance to drive businesses towards responsible AI practices. The Act introduces a preferential procurement system for companies that voluntarily undergo Human Rights Impact Assessments (HRIAs) for their AI systems. While this soft law approach encourages companies to act ethically, it lacks mandatory enforcement and stronger penalties for non-compliance. This difference in regulatory philosophy reflects a divergence in

how the two frameworks approach enforcement. The EU's model, with its high penalties, provides a strong deterrent against non-compliance, ensuring that businesses take regulatory requirements seriously. On the other hand, the AI Basic Act's emphasis on voluntary compliance might result in some businesses treating the law as a cost of doing business rather than a compliance obligation. Thus, strengthening enforcement mechanisms and clearly defining punitive measures for violations will be key to ensuring the Act's effectiveness [128].

Moreover, although the AI Basic Act provides a forward-thinking framework for AI governance, it faces challenges in adapting to the rapid evolution of AI technologies. The current legal framework may not be agile enough to address new, unforeseen risks that emerge with advancements in AI. For instance, technologies like generative AI, such as deepfake technology, could pose risks that are not adequately covered by the existing provisions. While the AI Basic Act mandates that businesses disclose when their AI-generated content is used, it does not provide sufficient regulatory clarity on how to manage the broader societal risks posed by such technologies. Therefore, the flexibility of the legal framework will be critical. To ensure that the AI Basic Act remains relevant, it will need to evolve with new technological developments, including through subordinate legislation and updates that address the emerging risks of generative and autonomous AI systems. Without the ability to quickly adapt to new technological realities, the law may become ineffective or obsolete, undermining its goal of fostering safe and ethical AI innovation.

2.2.2 AI Promotion Act

On May 28th 2025, Japan's Parliament approved the "Act on the Promotion of Research and Development and the Utilization of AI-Related Technologies" (Japan AI Promotion Act) [130], making Japan the second major economy after Korea in the Asia-Pacific (APAC) region to enact comprehensive AI legislation. Most provisions of the Act — except Chapters 3 and 4, and Articles 3 and 4 of its Supplementary Provisions — took effect on June 4th 2025, marking a significant transition from Japan's soft-law, guideline-based approach to AI governance to a formal legislative framework [40]. It is important to note that this legislative evolution builds upon Japan's established model of "agile governance", which emphasizes flexibility, adaptability, and multi-stakeholder participation to keep pace with rapid technological changes [22]. Rather than imposing binding obligations on private actors, this legislation is designed as a "basic law", similar in form and spirit to Japan's Science and Technology Basic Law or the Basic Act on Forming a Digital Society. The Act establishes strategic directions, policy guidelines, and national commitments aimed at promoting research, development, and deployment of artificial intelligence across all sectors.

The AI Promotion Act represents a significant step in embedding human rights considerations into the framework for AI development and governance in the country. Building on Japan's Social Principles of Human-Centred AI, the Act emphasizes that AI technologies must respect and uphold fundamental rights such as privacy, equality, and freedom of expression. This is reflected in its focus on transparency, accountability, and fairness, which are critical in mitigating risks like algorithmic bias, discrimination, and misuse of personal data. The Act's provisions also indirectly address

human rights by promoting privacy protections, ensuring data security, and fostering a fair and competitive environment for AI innovation. These measures aim to prevent harm to individuals, particularly in sensitive areas such as employment, healthcare, and law enforcement, where AI could significantly impact personal freedoms and opportunities. Furthermore, the Act's emphasis on inclusive governance ensures that all stakeholders, including citizens, private businesses, and local governments, have roles in shaping AI policies. Article 9 of the Act mandates "strengthened collaboration between the national government, local public entities, research institutions, and AI-utilizing businesses", while Article 15 requires the government to "promote public education and awareness of AI" to enhance citizens' ability to participate in policy debates. This participatory approach aligns with democratic principles and empowers individuals to have a voice in how AI technologies are implemented, reducing the risk of top-down or unaccountable decision-making. Notably, the Act avoids heavy-handed regulatory measures — such as mandatory algorithm audits for all AI systems — in favour of voluntary compliance supported by reputational incentives, a choice tailored to Japan's collaborative industrial culture. It encourages businesses to adopt self-regulatory frameworks and publicly disclose their AI governance practices, leveraging market pressure to drive accountability. For instance, companies that demonstrate strong human rights safeguards in AI may gain a competitive edge with consumers and investors increasingly focused on ethical tech. This approach fosters a "culture of responsibility" where human rights are not viewed as a regulatory burden but as a core component of sustainable AI innovation—critical for building public trust in technologies that are increasingly integrated into high-stakes areas like criminal justice and education.

Despite representing a milestone in the country's regulatory landscape for emerging technologies, the AI Promotion Act suffers from several shortcomings that may undermine its effectiveness. These issues primarily revolve around enforceability, oversight, and addressing systemic risks inherent in AI technologies. First, The Act's reliance on voluntary compliance, rather than legally binding obligations, creates significant gaps in accountability. Without punitive measures or mandatory oversight mechanisms, there is little recourse if businesses fail to adhere to ethical AI practices. For example, companies may prioritize profit over fairness or privacy, exacerbating risks like algorithmic bias or misuse of personal data. This lack of enforceability is particularly concerning in high-stakes applications, such as law enforcement or healthcare, where errors or abuses could have severe human rights implications [35]. Second, while the Act promotes privacy protections, it does not adequately address the risks associated with AI-driven surveillance technologies. For example, the use of facial recognition and other monitoring tools by public authorities or private entities could infringe on individual freedoms, particularly if deployed without meaningful oversight. This is especially relevant in Japan, where balancing technological innovation with privacy rights has been a longstanding challenge. The lack of clear safeguards against mass surveillance could lead to significant human rights violations, particularly in the absence of independent regulatory bodies. Third, the Act's focus on fostering innovation risks overlooking issues of equitable access to AI benefits. Populations in rural or underserved areas may face barriers to accessing AI-driven services, exacerbating existing social inequalities. Furthermore, without proactive measures to ensure inclusivity,

the benefits of AI may disproportionately favour well-resourced corporations or urban centres, leaving vulnerable groups further marginalized. Ultimately, the Japan AI Promotion Act seeks to strike a balance between promoting innovation and safeguarding human rights. By framing human rights as a core principle rather than a regulatory afterthought, the legislation aims to ensure that AI serves as a tool for social good, contributing to a society where technology empowers individuals without compromising their freedoms or dignity. However, its success will depend on how effectively these principles are translated into practice and whether the soft regulatory approach can adequately address the complex challenges posed by AI.

2.2.3 Provisions on the Administration of Algorithmic Recommendation in Internet Information Services

The Human Rights Action Plan of China (2021-2025) establishes the principle of "leveraging digital technologies to expand the space for the free and comprehensive development of all individuals" [59]. Rather than pursuing unified AI legislation at the outset, China has adopted a decentralized, scenario-specific regulatory approach, which facilitates a swift response to human rights issues arising during the early stages of artificial intelligence development.

As one of the first steps of this Plan, on December 31st 2021, the Provisions on the Administration of Algorithmic Recommendation in Internet Information Services was issued. This is the first departmental regulation in China and globally to specifically target algorithmic recommendation technology for regulation, marking a new phase of institutionalization and standardization in algorithm governance. Article 2 of the Provisions provide the first explicit definitions for five types of algorithms: generative synthesis, personalized recommendation, sorting and filtering, retrieval and filtration, and scheduling and decision-making. The Provisions are underpinned by higher-level legislation including the Cybersecurity Law, Data Security Law, and Personal Information Protection Law, focusing on issues such as illegal information, algorithmic discrimination, personal information protection, and safeguarding minors. Furthermore, they establish a preliminary algorithm governance framework encompassing pre-emptive prevention, in-process compliance, and post-incident redress, which embodies the "protect, respect and remedy" principle of the UN Guiding Principles on Business and Human Rights.

In implementing the State's duty of protection, Article 6(2) of the Provisions adopts a negative approach by stipulating that algorithmic recommendation service providers shall not use such services to infringe upon the legitimate rights of other users. Specifically, the Provisions establish a system for algorithmic filing and classification-based management.The Provisions require providers of algorithmic recommendation services possessing public opinion attributes or social mobilization capabilities to fulfil their registration obligations. This tiered classification approach shares common ground with the European Union's Artificial Intelligence Act in regulating high-risk AI systems, which reflects the country has taken into account the diversity and impact differences of algorithm technology when fulfilling its protection obligations. Additionally, the Provisions place particular emphasis on the potential impact of algorithms on vulnerable groups such as minors, the elderly, and workers, requiring algorithm service providers

to implement effective measures to safeguard their lawful rights and interests. For instance, addressing the issue of minors becoming addicted to online activities, the Provisions stipulate that algorithmic models must not be designed to induce such behaviour.

As developers, operators and users of artificial intelligence technology, algorithmic recommendation service providers possess not only first-hand knowledge of technological advancements but also direct engagement with users, granting them a governance advantage over governmental bodies development of all individuals" [153]. Consequently, they should exercise self-regulatory oversight. The Provisions translate corporate responsibilities regarding respect for human rights into specific, actionable compliance obligations, providing algorithmic service providers with clear guidance. One is the right to be informed of algorithm information, which means informing users of the relevant information recommended by the algorithm, and disclosing the basic principles, target intentions, and main operating methods of its services. However, this transparency requirement lacks a tiered approach. For algorithms of high technical complexity and commercial sensitivity, excessive disclosure may compromise corporate trade secrets and core competitiveness; conversely, for low-risk algorithms, existing transparency requirements may prove insufficient to safeguard users' right to know. This one-size-fits-all transparency mandate struggles to strike a balance in practice, potentially leading enterprises to opt for symbolic disclosure rather than substantive transparency. Secondly, the selection right of algorithms is stipulated, which provides users with choices that do not target their own personality, or facilitates the closure of algorithm recommendation services. Thirdly, the Provisions require algorithm recommendation service providers to fulfil their primary responsibility for algorithmic safety. They must establish and improve management systems and technical measures covering algorithmic mechanism review, technological ethics assessment, user registration, information publication review, data security and personal information protection, anti-telecommunications network fraud, security evaluation and monitoring, and emergency response to security incidents. Providers shall formulate and publicly disclose relevant rules governing algorithm recommendation services.

Moreover, beyond judicial remedies based on national jurisdiction, the Provisions also stipulates non-judicial redress mechanisms. Article 22 stipulates that providers of algorithmic recommendation services shall establish convenient and effective channels for user appeals and public complaints or reports, clearly define processing procedures and response time-frames, and promptly receive, handle and provide feedback on the outcomes of such cases. This non-judicial redress mechanism provides users with a relatively low-cost avenue for relief, facilitating the prompt handling of complaints and enabling direct redress.

2.2.4 ASEAN Guide on AI Governance and Ethics

At the 4th Digital Ministers Meeting of the Association of Southeast Asian Nations (ASEAN), which was held in Singapore, the Guide on AI Governance and Ethics (AI Guide) has been released. It serves as a practical guide for organizations in the region that wish to design, develop, and deploy traditional AI technologies in commercial and non-military or dual use applications. AI Guide highlights seven principles, including transparency, fairness, security, reliability, human-centricity, privacy and accountabil-

ity. When it comes to human-centricity, AI Guide proposes that:

"It is key to ensure that people benefit from AI design, development, and deployment while being protected from potential harms. AI system should be used to promote human well-being and ensure benefit for all. Especially in instances where AI systems are used to make decisions about humans or aid them, it is imperative that these systems are designed with human benefit in mind and do not take advantage of vulnerable individuals" [148].

AI Guide advances the ASEAN Digital Masterplan 2025's Desired Outcome (2.7) which is to adopt regional policy to deliver best practice guidance on AI governance and ethics, IoT Spectrum and technology. The Guide is explicitly structured to be a "living document", allowing it to evolve in response to emerging technologies and governance challenges [81].

Recognizing the rising prominence of generative AI, ASEAN released an Expanded ASEAN Guide on AI Governance and Ethics — Generative AI in 2025 to supplement and support AI Guide with policy considerations related to generative AI. It outlines six core risks, including mistakes and anthropomorphism, inaccurate content and disinformation, deepfakes, impersonation and malicious use, IP rights infringement, privacy breaches and biased outputs.

The Guide's principles overlap with rights recognized under international human rights law, such as right to information, access to remedy, right to non-discrimination and equality, right to privacy and human dignity. Compared to mandatory measures, ASEAN's preference for non-binding, consensus-driven, and flexible frameworks, which may be more politically feasible in a region with diverse governance models but weaker in enforceability for rights protection. The Guide represents an important regional step toward ethical AI governance, but its integration of human rights is indirect, implicit, and aspirational rather than binding.

2.2.5 Future Trends of Integrating Human Rights into AI Governance in Asia

Currently in Asia, developments in AI governance suggest that, in the future, Asian countries are more likely to opt for soft-law instruments like regulations as opposed to hard-law mechanisms like legally enforceable constitutional rules. As mentioned in table 5 several Asian countries have been working on voluntary guidelines have been introduced to shape an AI regulatory framework which incorporates human rights considerations. This trend reflects both pragmatic and structural factors. On the one hand, voluntary guidelines allow governments to respond quickly to technological changes without the political and legal complexities of passing binding legislation. On the other hand, voluntary guidelines play a strategic role in shaping industry norms and preparing the ground for future regulation. They encourage companies and research institutions to internalize human rights principles while leaving space for innovation and experimentation. Over time, these soft frameworks may crystallize into standards that influence binding lawmaking.

Even where human rights are not expressly stated as a consideration, Asian countries have been impressing the need to regulate AI in order to protect human interests. For example, Singapore has developed sector-specific guidelines such as the Artificial Intelligence in Healthcare Guidelines, which provide detailed guidance for the design,

development, and deployment of AI medical devices, based on the principles of fairness, responsibility, transparency, explainability and patient centricity. Another example is China, where the Robotics + Application Action Implementation Plan not just promotes AI integration in healthcare, elderly care, and education, but has asserted that it will also look into best practices and risk management so as to "develop a culture of responsible AI development" [121, 71].

In summary, while Asian countries are converging on the need to embed human rights into AI governance, their approaches remain heterogeneous and often less enforceable than European frameworks. South Korea's AI Basic Act represents a comprehensive and rights-oriented framework, yet its centralised governance model and relatively weak penalties may undermine effective compliance. Japan's AI Promotion Act embeds strong principles of inclusivity and fairness but relies on voluntary compliance and lacks a risk-classification system, creating accountability gaps in high-stakes areas such as surveillance. China's scenario-specific regulations mandate corporate duties to respect human rights but risk fragmentation across agencies and encourage symbolic disclosure rather than substantive transparency.

At the regional level, a human rights-based approach to AI regulation is gaining momentum. However, these regulations are mainly soft law in nature, where enforceability, and liability for unethical actions, is impaired. For example, while the ASEAN Guide on AI Governance and Ethics is a step in the right direction, it remains a guide where adherence is voluntary. Another example is the Asian Forum on Human Rights that took place in China, where participants unanimously agreed that technology must be fundamentally oriented towards the protection of human rights [93, 111]. Once again, while the intention is encouraging, more needs to be done with regards to monitoring and accountability.

Taken together, Asia's frameworks prioritise flexibility, innovation, and state-led development, in contrast to Europe's binding and rights-based approach. Whereas European instruments treat algorithmic bias and discrimination as direct human rights violations, Asian systems more often frame them as technical or governance challenges to be managed. This divergence underscores the importance of Asia–Europe dialogue: Europe can contribute enforceable rights safeguards, while Asia offers models of regulatory adaptability and innovation.

2.3 Human Rights and AI at the regional level: Europe

As artificial intelligence has become increasingly embedded within the social fabric, the European Union has concentrated its efforts on developing governance and regulatory frameworks for these technologies. Such an initiative stems from the intention to find a balance between fostering technological research and development at the Union level and upholding principles and values central to European legislation, such as respect for fundamental human rights, consumer protection, fair competition, and the rule of law. These efforts have given rise to a set of harmonized European strategies aimed at reconciling the interests of the technology industry and related companies with the safeguarding of end users.

Among these strategies, the following sections will examine in greater depth: the General Data Protection Regulation (Section 2.3.1), the Framework Convention on Ar-

tificial Intelligence and Human Rights, Democracy and the Rule of Law (Section 2.3.2), the Human Rights, Democracy, and Rule of Law Impact Assessment Methodology (Section 2.3.3), and the Artificial Intelligence Act (Section 2.3.4).

2.3.1 General Data Protection Regulation

Among the issues that have emerged as particularly pressing within the European landscape — initially with the advent of the internet and subsequently with the progressive development and consolidation of artificial intelligence technologies — the protection of privacy and personal data occupies a position of primary importance. Within the European legal order, these principles are recognised as essential components of the corollary right to personal identity and individual autonomy. However, once the imperative of safeguarding which pertains most intimately to each subject of law — including, among others, data relating to their person, health, habits, and lifestyle — shifted to the digital domain, the legal framework was confronted with a conceptualization of such notions, as well as with related management and protection strategies, that proved to be arguably incomplete and potentially ineffective [144, 1].

Originally, the right to data protection was considered an aspect of the right to privacy. Consequently, the safeguarding of data pertaining to the personal sphere of each data subject *per se* emerged as a subsequent concern, albeit its subsequent rapid dissemination [52]. The first Data Protection Act was passed in 1970 by the German government, and it was followed by similar legislation in Sweden (1973), Australia (1978), Norway (1978), France (1998), and the United Kingdom (1998). It was not until the advent of the Charter of Fundamental Rights of the European Union (2009) that the right to the protection of personal data was officially recognized as a fundamental right, at least within the European context [164].

This approach was then consolidated by the General Data Protection Regulation (GDPR), which came into force in 2018, becoming an enforceable and binding legislative instrument for member states [141, 169]. The framework of this act is based on the principles expressed in the EU Data Protection Directive of 1995, and then establishes new rights in the field of privacy and data protection, such as data portability and the right to be forgotten. This legislation reaffirms Europe's unique approach in establishing data protection as a fundamental right. Nevertheless, according to some scholars, this approach is incongruent with numerous business practices in the digital age [27]. The core of this issue lies in the fundamental nature of human rights, which are inherently non-tradeable. Consequently, the pricing of data, whether for exchange, sale, or transfer, should be prohibited. However, it is evident that this practice is prevalent and frequently justified by the institute of informed consent — which is becoming increasingly problematic in the context of AI and digital exchange of information. In fact, the GDPR requires that the data subject must give their free, specific, and unambiguous consent for their data to be processed legally — as for Article 4 of the document. Nonetheless, it is important to note that such agreement can be withdrawn at any time, as fundamental rights are inalienable. This observation underscores what some scholars have characterized as an ambiguity in the approach adopted by the European legislator. On the one hand, Europe attempts to circumvent the commodification of personal data, a position that is consistent with the notion that human rights should not be regarded

as a commodity to be traded in order to fuel market growth or advance technological development [28]. On the other hand, it is clear that the European Union has expressed a clear intention to develop a data-driven economy internally, thus making GDPR compliance complex for both businesses and European institutions themselves [105, 4].

Even if the GDPR is often discussed in compliance and technical terms, at its core it was designed as a human rights protection instruments, as expressed in Recital 1 of the act. Among the rights that are primarily highlighted there are: the *right to information* — protected by Articles 13 and 14, which aim to enhance transparency on how data is collected, the legal basis for collection, and how long and for what purpose it is retained — the *right to access* and the one to *restricted processing* — protected under Articles 15 and 18 respectively, which govern an individual's right to know whether and how their data has been processed and to manage safeguards in the event that the legitimacy of its use is disputed.

Due to its focus on rights protection, the GDPR was also the first piece of legislation to focus on the concept of 'risk to the rights and freedoms of natural persons' — as for Articles 24, 25, 32, 35 — thus paving the way for a long series of regulatory and governance documents based on a risk-based approach [54]. In particular, in Recital 75 and Articles 24 and 25, such a risk is considered as the combination of probability and severity of physical, material, or non-material damages resulting from data processing.

In the document, this approach is reflected in the outline of a Data Protection Impact Assessment (DPIA), as for Article 34. It has been conceived as a European version of the Privacy Impact Assessment (PIA) which had been developed by the OECD and practically applied in legal frameworks in Canada and Australia. The purpose of the DPIA in the GDPR is to ascertain the existence of risks to the rights data holders (Recital 1(2) and Recital 75). This wording provides a broad spectrum of flexibility, extending beyond privacy protection to also cover the rights to dignity, freedom of expression, non-discrimination, and access to services. Specifically, this tool is intended to verify the proportionality and necessity of data processing operations, to highlight potential associated risks, and to provide evidence both of compliance with data protection principles and of negligence, which can then be linked to the attribution of responsibility for any infringements that have occurred [16]. Following the procedure specified in Article 37 of the GDPR, the European legislator aimed to structure an assessment that, based on a risk-based logic, could serve both as a foundation for legal accountability and as a means to prevent or mitigate foreseeable risks. Indeed, upon identifying potential risks in the use of a particular AI technology, the indication of countermeasures to be applied in order to limit them is required. However, if risks persist, Article 36 mandates that controllers consult the supervisory authority and inform them of the issue. This mechanism establishes an ex ante dialogue that strengthens regulatory guidance and is intended to facilitate regulatory compliance.

In light of the above, it is important to emphasize that the DPIA required by data protection regulation is controller-driven. This entails that it is conducted by the controllers themselves, without the mandatory review by external supervisory bodies — unless residual risks remain, which must be significant enough to necessitate reporting. Moreover, risk assessment often follows a "tick-box" exercise logic, limited to yes-or-no answers, without deep analysis or a need for detailed problematization of the results obtained [99]. Additionally, despite the fact that risk assessment variables for

AI systems have been somewhat inaugurated by this legislation, the document does not clearly define thresholds for categorizing risks as high, medium, or limited [32]. This evaluation is left to the interpretation of internal company personnel conducting the assessment, which leaves room for legal ambiguity regarding the outcomes obtained. Furthermore, many DPIAs are not publicly disclosed, and competent authorities are not obliged to review them unless damage occurs or irregularities are reported [32, 99]. These factors undoubtedly raise questions about the concrete effectiveness of the safeguards provided by the GDPR in protecting fundamental human rights — foremost among them the right to privacy — despite the regulation's significant theoretical and conceptual influence.

2.3.2 Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law

On September 5th 2024, the Council of Europe adopted the *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, which establishes a monitoring mechanism to ensure the protection of fundamental human rights throughout the entire lifecycle of AI systems. This document constitutes the world's first binding international treaty specifically dedicated to addressing the human rights implications of the development and deployment of AI technologies [143].

By virtue of its legal status, the Framework Convention obliges signatory states to incorporate its principles and provisions into national legislation and administrative procedures, thereby embedding human rights safeguards within domestic AI governance frameworks.

This initiative also reflects a broader and increasingly consolidated trend within recent European regulatory efforts: the commitment to multi-stakeholder engagement. In this context, the drafting process included the participation of representatives from academia and industry, as well as international organizations and civil society actors at the global level.

At its core, the Framework Convention sets out a series of fundamental principles that must guide state action in the development and use of AI technologies. Among these, the protection of human dignity is explicitly emphasised. Often described as a "constellation right" [176], human dignity is considered by the majority of legal scholars as the foundation upon which the recognition and enforcement of all other fundamental rights depend [127]. Closely related is the principle of individual autonomy, understood as an expression of self-determination and freedom from external influence or coercion [100]. Furthermore, the act incorporates considerations related to the environmental sustainability of AI systems, emphasizing the need for risk mitigation strategies in the deployment of technologies with potentially significant environmental impact [143]. It also reaffirms the principle of equality and non-discrimination, with particular attention to the protection of vulnerable groups — which is a recurring concern in European policy frameworks [134]. Great attention is focused on the theme of algorithmic bias and its potential to undermine individuals' human rights, especially in contexts where automated or algorithm-supported decisions significantly affect people's lives. Naturally, ensuring the protection of fundamental rights is inseparable from the principles of accountability — requiring a clear allocation of responsibilities for

harms resulting from rights violations — and human oversight, which the Framework Convention identifies as a critical safeguard in the design and deployment of AI systems.

Based on the provisions of the treaty, signatory states are required to collect relevant information regarding the use and characteristics of AI systems deployed within their national territory and to make this information accessible to the populations potentially affected by them. Such information must be presented in a manner that is both understandable and usable by the public, enabling individuals to make informed decisions about whether to rely on AI systems and to assess the trustworthiness of AI-supported outcomes.

The *Conference of the Parties,* established by the Framework Convention itself and composed of representatives from the signatory states, is the body tasked with supervising and monitoring the implementation of the act's provisions. Its mandate also includes facilitating stakeholder consultations to support understanding and address challenges encountered in the implementation process. National authorities are further empowered to introduce red lines—that is, outright bans—on specific AI systems deemed potentially harmful to the protection of fundamental rights. Such determinations are made through the application of a risk-based approach, designed to assess the impact of AI systems on human rights, democracy, and the rule of law. Rather than relying on a fixed classification of AI systems into predefined risk categories, the treaty promotes a system-specific and iterative assessment throughout the lifecycle of the AI technology in question. This tailored evaluation is intended to ensure contextual adaptability in risk management.

Despite the undoubtedly revolutionary scope of this instrument, the treaty is enforceable only within the jurisdictions of the states that have ratified it. As such, non-signatory states are under no legal obligation to comply with its provisions. While this does not diminish the importance of the issues addressed by the Framework Convention, it does limit its potential impact. This limitation once again brings to the forefront the challenge of the "territoriality" of AI regulation and, consequently, the difficulty of ensuring consistent protection of fundamental human rights in the global development and deployment of AI.

2.3.3 Human Rights, Democracy, and Rule of Law Impact Assessment Methodology

Among the initiatives promoted by the Council of Europe, great expectations are cultivated for the adoption in November 2024 of the Human Rights, Democracy, and Rule of Law Impact Assessment Methodology (HUDERIA), which aims to evaluate the impact of AI systems on fundamental rights, democratic governance, and the rule of law [85]. This methodology is designed for use by both public and private sectors to support them in identifying, managing, and mitigating risks associated with the specific AI technology under review, throughout its entire lifecycle. Therefore, this framework is intended to be iterative and not a one-time assessment. The perspective adopted by such a new form of impact assessment tool is socio-technical, meaning it takes into account the fact that each AI system is or is intended to be embedded in a social context influenced by cultural, legal, and economic factors, and is affected by human decisions [92].

Based on these underlying assumptions, HUDERIA is structured into the following phases:

- **Context-Based Risk Analysis:** This phase analyses the specific context in which the AI system in question was developed and is used, bringing to light potential risk elements concerning the safeguarding of fundamental rights, democratic values, and the rule of law.
- **Stakeholder Engagement Process:** The involvement of all the most relevant stakeholders is considered essential to evaluate the perspectives of those who may be directly or indirectly impacted and to ensure effective transparency.
- **Risk and Impact Assessment:** The possible risks identified are evaluated based on criteria such as likelihood, severity, and potential impact. This also helps assess whether the AI system under examination is appropriate for its intended purpose and how it might affect the enjoyment of fundamental rights by users.
- **Mitigation Plan**: The final stage aims to determine which measures can be implemented and which regulations may apply to limit risks and protect potentially affected parties. This phase is designed to be iterative and involves monitoring the AI system throughout its operational lifespan.

Such a methodology is not intended to be a legally binding instrument. Rather it serves to complement the Council of Europe's Framework Convention on Artificial Intelligence and other legal documents that establish a clear framework of norms without specifying concrete application tools for the rules imposed — such as in the case of the AI Act 2.3.4. The added value of this initiative lies in setting a precedent and model for a rights-based, anticipatory regulatory approach to AI that embraces democratic principles while going beyond traditional focus on mere formal and technical compliance. In fact, its principal merit is the attempt to incorporate into the impact assessment procedure evaluations that consider the relevant social context and the related social and cultural dynamics that are involved in the deployment of new technologies.

2.3.4 Artificial Intelligence Act

The Artificial Intelligence Act (AI Act) represents the first attempt to establish a uniform and systematic regulation of artificial intelligence. It is the result of a long debate between political parties, which also involved experts in various fields of research related to the development and impact of new technologies, representatives of users of these technologies, and industry representatives — including programmers and designers [19]. The objective was to structure a regulatory framework that would not hinder the development of AI in Europe, but would also guarantee the protection of citizens from potential harms, with particular regard to those which could affect fundamental human rights [80]. In this regard, one of the aspirations of the European legislator was that this act would generate the so-called 'Brussels Effect', as had already happened in the past with the regulation regarding data and privacy protection and management. In fact, there is a widespread belief that drawing inspiration from *legislation made in Europe* on the development and commercialisation of AI could also promote the spread

of technologies that, by design, incorporate high standards of protection of essential rights such as dignity, non-discrimination, and respect for the rule of law and democratic principles.

Thus, the AI Act, signed in its final version in June 2024, took effect in August 2024, but numerous steps are still necessary for it to become operational and effective.

It establishes a classification of AI systems based on the risk of harm to which their users are exposed. Systems that expose people to an unacceptable risk to their physical or mental safety and to their fundamental rights are considered prohibited technologies and addressed in Article 5 and Annex III. These are followed by high-risk systems, as defined in Article 6 of the Act, which have the potential to inflict significant harms and are subject to stringent design and operational requirements. Technologies in the limited risk category are those that could expose users to risks of lack of transparency, and for which there is an obligation to make the user explicitly aware that they are interacting with an artificial agent or that the outcome produced is the result of AI. Similar attention to the need to disclose when interacting with an AI rather than a human operator or when exposed to AI generated content is also strengthened for general purpose AI (GPAI), addressed separately in Title VIII. In this last case, it is important to make it clear to the user also the ability of the system to autonomously generate its outcomes, and the datasets and training methods used, including their limitations. All other systems are classified as minimal risk and are essentially unregulated.

The risks that AI Act considers are those affecting the fundamental rights of individuals, with particular emphasis placed on human dignity, privacy, non-discrimination, and equality. Therefore, even though the rules regulating the requirements for specific classes of AI technologies and the necessary risk mitigation measures bear similarities to those found in safety regulations, the Act is conceived by the European legislator as a human rights protection instrument.

In line with this goal, Article 27 mandates that a Fundamental Rights Impact Assessment (FRIA) be conducted prior to the deployment of a high-risk AI system [129]. This obligation applies to public bodies and private entities offering public services, as well as deployers under Annex III point 5(b) and (c) — e.g. companies using AI for credit scoring, or for risk analysis purposes in life or health insurance. This assessment must be based on a description of the modalities in which the system is expected to be used — including timing and frequency — a disclosure of the categories of individuals who may be affected by use of the technologies under analysis, and the harms that might potentially occur. Furthermore, a mandatory description of the strategy to be implemented if the anticipated risks materialize must be included, together with the exact type of human oversight intended for deployment. All the above information must be collected before first use of the AI system. Subsequent deployers may rely on previously completed assessments unless they deem it appropriate to update the details due to obsolescence or initial lack of accuracy. Once this assessment has been completed, market surveillance authorities have the power to permit the placing on the market of systems that have successfully passed the FRIA.

Nevertheless, Article 46(1) introduces a significant exception to the procedure outlined above, potentially undermining the protective scope of Article 27 regarding human rights. In fact, Article 46 allows the use of high-risk AI without the need to demonstrate no or marginal impact on fundamental rights for reasons such as public security

or environmental protection. Although the provision refers to "justified reason" — suggesting that the exceptional circumstances which lead to the application of this norm cannot be entirely arbitrary — it is true that this expression is in itself a source of ambiguity in legal terms. Indeed, it raises doubts about the actual impact of the FRIA obligation in the AI Act [167]. Also the European Data Protection Supervisor has expressed apprehensions regarding the exemptions that could endanger the safeguards for high-risk AI systems. The primary source of concern for this European body is the potential inconsistent application of protections designed to guarantee fundamental rights of individuals, leading to legal uncertainty in one of the core regulatory fields [156].

2.3.5 Open Challenges in the European AI Strategies

The previous sections present some of the most significant examples of strategies and interventions implemented by European institutions to address the risk of human rights infringements perpetrated by or caused through AI systems. As can be observed from the overview, the approach favoured by Europe is the risk-based approach. This choice is likely guided by the fact that one of the main objectives pursued by member states in regulating new technologies has been not only to ensure the protection of end users, but also to foster technological development in this region of the world.

To achieve this, a framework based on the logic of product safety regulation is conducive to greater legal certainty for providers. Indeed, providers already have prior experience with compliance and consequences under safety laws, and therefore may be more inclined to conform to procedures and rules similar to those with which they are already familiar [67]. However, product safety norms are primarily designed to address health and safety risks [133]. Consequently, while they contribute to creating safer products, they only partially cover the range of risks potentially imposed on fundamental rights by AI technologies, especially high-risk ones.

The main challenge is that the concept of risk and that of human rights — particularly fundamental rights — belong to legal, evaluative, and thematic categories that are fundamentally misaligned. On the one side, a fundamental right is often regarded as an inalienable attribute of every human being, by virtue of our shared humanity [65, 64]. It is a right to which not even the holder can lawfully renounce. On the other side, risk is situated within the domain of verifiability, quantification, and systematic analytical processes. Risk must be measured so it can be anticipated, mitigated or eliminated altogether. In many cases, risk may be considered tolerable, when weighed against other contextual factors, whereby trade-off among likelihood, severity, and potential benefits are made. In other words, one can establish a threshold beyond which risk is deemed intolerable, and below which it may be considered acceptable for pragmatic reasons.

Such thresholds, however, cannot be applied to fundamental rights. Legal theory and judicial practice recognize no threshold within which the impairment of a human right may be considered tolerable and beyond which sanctions are triggered. To do so would undermine the very notions of essentiality and inviolability that define these rights.

What is possible — and indeed common — is the balancing of multiple rights, guided by rigorous precedents and norms, undertaken on a case-by-case basis, often by apex courts such as the European Court of Human Rights or the International Court

of Justice. Such balancing exercises lie beyond the actuarial logic of risk, a logic more commonly associated with business strategy, civil liability frameworks, and corporate policy, and not with the protection of core human values [6].

Scholars seeking to reconcile what appears an irreducible conceptual and applicative gap have frequently resorted to the use of proxies – indirect measures or surrogate categories of risk — intending to avoid clear-cut quantitative frameworks [95]. Nonetheless, such strategies rely on assessments conducted by individuals who, even if competent, may lack complete impartiality or may not guarantee consistency in judgment and implementation of countermeasures. Mechanisms aimed at limiting the arbitrariness of possible interpretations of certain norms are sometimes embedded within the regulations themselves, as in the AI Act — specifically, Article 42 [129]. Nevertheless, even there, reliance is placed on external actors, such as in the case of external certifications and the prominent reference to harmonised standards. These latter instruments, in particular, present non-negligible challenges.

By definition, harmonised standards are technical tools drafted predominantly by large corporations and through processes that are largely removed from the democratic procedures normally employed in the creation of new regulations or laws. Furthermore, notified bodies are often composed mainly of technical experts who typically lack experience with the complex mechanisms involved in recognising and protecting fundamental human rights. Indeed, establishing an effective defence of these rights requires the ability to interpret concepts such as 'interference with human rights' or 'risk to fundamental rights', which are highly complex and multifaceted notions that often challenge even jurists specialized in international law. In fact, human rights are inherently matters of policy and legal balancing, difficult to quantify and systematise comparably to more tangible risks commonly assessed by standards [166] — such as those arising from the use of specific chemical agents or adherence to particular corporate procedures rather than others.

In light of these considerations, a rights-based logic and a more holistic approach to the impact that AI systems may have on individuals and society as a whole could prove most effective, at least in regulatory and governance frameworks that prioritise the protection of fundamental rights. From this perspective, the initiatives and outlook adopted by the Council of Europe appear to outline a more effective strategy for safeguarding human rights in the digital era. The Council notably includes an evaluation of the context in which AI technologies are used, paving the way for a better assessment of collective and systemic repercussions of technological development — rather than merely individual ones. Moreover, the range of rights covered by this approach is broader and more flexibly expandable. The impact assessment methodology proposed by the HUDERIA (2.3.3) is also designed as an iterative process, ensuring coverage throughout the entire lifecycle of AI systems while adapting to any acquired capabilities, ongoing learning processes, or technological upgrades. In doing so, it may adequately address the persistent challenge posed by the mismatch between the rapid pace of technical development and the rhythms of legal adaptation.

2.3.6 Future Perspectives

The European Union is resolutely committed to becoming a global leader in the responsible development of artificial intelligence. To this end, on April 9, 2025, the so-called AI Continent Action Plan was launched, aiming to transform Europe into an "AI Continent" grounded in the principles of transparency, trust, and respect for democratic values [25]. A human-centric approach stands as the cornerstone of this agenda, underpinning a suite of measures intended to improve data access and promote the responsible advancement and deployment of AI systems and AI-driven solutions across key sectors such as industry, sustainability, education, and healthcare.

To realise these ambitions, the AI Continent Action Plan sets out a comprehensive roadmap to ensure the safe and fundamental rights-respecting deployment and market diffusion of AI technologies. Central to this effort is the progressively closer and interdisciplinary cooperation between the economic sector, technical experts, and policy makers, fostering a successful intertwining of technological excellence with ethical leadership [113].

Naturally, achieving these objectives necessitates maintaining a global outlook — one that accounts for the technologies being developed, the emergent technological needs, and the evolving legislative frameworks in other regions of the world. The effort to establish an efficient and human rights-compliant AI continent cannot succeed without open collaboration and dialogue among leading economic actors in the global market, nor without the planning of a governance and regulatory framework that is as harmonized as possible. This is particularly critical given that AI, in light of its technical and operational characteristics, tends to transcend both physical and legal national boundaries. Thus, an overly fragmented AI policy approach would merely incentivise the development of technology in under-regulated or more business-friendly jurisdictions, leaving fundamental human rights at risk.

3 Thematic focus

Among the fundamental rights implicated by the development, deployment, and dissemination of intelligent systems, certain rights have garnered particular attention from international policymakers. Chief among these is the right to privacy, whose protection has become increasingly complex in light of the pervasive and extensive use of data intrinsically linked to identifiable individuals throughout the technological lifecycle. Consequently, another category of rights that has attracted considerable focus within global governance concerns equality and non-discrimination. In fact, progressive automation of AI systems and their expanding role in various decision-making processes expose them to risks of unfair treatment, bias, and the perpetuation of social inequalities embedded in the underlying data. Faced with these two examples of fundamental rights potentially endanger by AI, a crucial challenge that different states around the world are called upon to address is that of guaranteeing access to the judicial system and remedies for those adversely affected. As discussed in section 2.3.5, a right is inalienable not only when universally recognized as a fundamental entitlement,

but also when the legal framework enables rights holders to enforce it at any time and against any infringers. Thus, safeguarding the remedial system is essential to upholding the rule of law and securing substantive, rather than merely formal, equality before the law.

Therefore, the following sections aim to explore these three central themes regarding the approach that ASEM countries adopt to the protection of human rights in AI: (i) the right to privacy and data protection, (ii) the right to equality and non-discrimination, and (iii) the access to justice.

3.1 Privacy and Data Protection

The right to privacy is protected by Article 12 of the Universal Declaration of Human Rights. In particular, it stipulates that "no one shall be subject to arbitrary interference with his privacy, family, home, or correspondence, nor to attacks upon their honour and reputation" [11]. The International Covenant on Civil and Political Rights (ICCPR) provides protection against arbitrary or unlawful interference and attacks on privacy in Article 17. Confirming the fact that this right is highly personal and fundamental, dependent not on additional factors but solely on belonging to the humankind, the Convention on the Rights of the Child also recognises it in Article 16 for minors, even if they do not yet have full legal capacity.

The advent of the internet, artificial intelligence, and related technologies has quickly highlighted the practical difficulty of guaranteeing such a right on a large scale. This is linked to the role that data plays in such a landscape of innovation and development, especially when it comes to data that pertains to the private and personal sphere of the individuals who hold it.

In such an already complex and dynamic scenario, a further difficulty in protecting the rights analysed here also stems from the fact that privacy and data protection are not conceptualised in the same way and do not have the same relevance everywhere in the world. Examples of this divergence are evident in the ASEM countries too, resulting in different levels of attention and urgency in their regulation and governance. Europe, in fact, considers privacy — and, since the adoption of the GDPR, data protection as well (Section 2.3.1) — to be fundamental rights, protected by a unified and enforceable legal regime. As a corollary to the protection thus guaranteed to these rights, there is a particular emphasis on the theme of individual autonomy, individual control over what belongs to the subject, and consent as an essential institution for allowing others access to this sphere of identity and subjectivity. At an even broader level, this vision can be traced back to property as a central right in European legal systems — but we could also say Western legal systems — since Roman law [86, 61]. Asian countries, on the other hand, show a more heterogeneous approach, which includes both the adoption of forms of strict law for the protection of privacy in some states, and greater attention to the delicate balance between privacy, state control, and collective economic growth in other states [60]. In fact, in Asia, the concept of privacy is often intertwined with economic development priorities, collective social values, and government roles that push for a more peculiar balance between state interests and individual rights than what has historically been observed in Europe.

Given such a multifaceted approach to data protection, it is nevertheless worth noting that the challenges facing policymakers are similar on an international scale, due to the comparable technical requirements of the intelligent systems for which these data are essential.

3.1.1 Sources of Privacy Harms in AI Systems

The data collected, analysed, and stored to enable the functioning of AI systems often concern highly personal aspects of individuals' lives, such as health, habits, location, political or religious beliefs, or sexual orientation. Because AI technologies depend so heavily on this type of information, they create significant risks for the protection of privacy and personal data in the digital age.

These risks can be better understood by distinguishing between two categories of privacy harms: those linked to how data are collected and managed, and those linked to how data are subsequently used by AI systems. The first category includes vulnerabilities in gathering, storing, and re-using personal information, often without clear consent or adequate safeguards. The second arises when AI systems process or infer new information from data, for instance by profiling, biometric analysis, or large-scale surveillance.

Taken together, these challenges illustrate the continuing difficulty of reconciling the functioning of AI with the protection of privacy. Despite ongoing regulatory initiatives, high risks to fundamental rights remain and require sustained technical, legal, and institutional responses.

Sources of Privacy Harms in AI Systems

Privacy harms from AI generally arise from two categories:
(1) Data collection and management

- **Excessive data gathering:** collection beyond what is necessary [94].
- **Ambiguous consent:** unclear or overly complex authorisation mechanisms [50].
- **Weak anonymisation:** anonymised data can be re-identified [53].
- **Data re-use:** information repurposed without consent [83, 125].
- **Regulatory mismatches:** conflicting legal regimes in cross-border contexts [82, 15].

(2) Data use by AI systems

- **Opaque governance:** unclear how data drive algorithmic outcomes [31, 112].
- **Profiling:** detailed user profiles built for decision-making [147, 96].
- **Algorithmic training:** models trained on personal data without consent [168].
- **Mass surveillance:** large-scale monitoring of individuals [98].
- **Biometric analysis:** use of facial or bodily data with chilling effects [39, 75, 155, 97, 102].

3.1.2 Illustrative Examples

The ASEM region has seen multiple cases where the use and processing of personal data by AI systems has caused harm. Risks from AI-enabled data collection, profiling, and inference often stem from scale, opacity, and function creep. A central concern is the growing gap between what data subjects expect and how their information is used, coupled with the difficulty of contesting inferences drawn from data they never explicitly provided. See Appendix C.1 for the full set of illustrative cases and citations. Despite sustained attention to privacy in global governance, effective data protection in the digital age remains unresolved.

Cases from ASEM countries also reveal uneven visibility: more have been documented in Europe than in Asia, reflecting differences in transparency and enforcement. While this indicates growing attention by regulators, it also underscores the limits of existing frameworks in preventing and addressing violations of fundamental rights.

3.1.3 Legal and Policy responses

As already highlighted in the previous sections, the issue of privacy has been one of the most pressing concerns for researchers and policymakers with the advent and progressive spread of intelligent systems. Such a focus is naturally due to the awareness that AI requires data in order to be developed, trained and, ultimately, to function. This has brought to light what could in some ways be considered an irreducible aporia between (i) the need to protect information that draws on the private and highly personal sphere of legal subjects and (ii) the need to make as many data as possible available to AI, in order to enable its effective integration into civil society [1].

Therefore, ASEM countries have developed multiple regulatory attempts that aim to balance support for technological development with the goal of protecting fundamental human rights. Among these attempts, the GDPR — which was discussed in detail in section 2.3.1 — certainly stands out. Through it, Europe has attempted to outline a replicable framework model that guarantees rights such as data portability, access, and erasure. In parallel, the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data — also known as European Convention 108 — in its revised 2018 version, constitutes the only multilateral agreement on personal data protection with binding legal force at the global level [24]. It has played a primary role in reinforcing obligations related to algorithmic accountability and breach notification. Naturally, the AI Act also aims to safeguard privacy as one of the fundamental human rights that regulation seeks to protect. While the Act leaves intact the provisions of legislation directly governing the fair processing of personal data, it repeatedly imposes stringent requirements concerning transparency and the significance of human oversight, particularly when high-risk systems are involved [129].

With respect to the Asian region, in 2024 the Association of Southeast Asian Nations adopted a *Guide on AI Governance and Ethics*, highlighting the importance of transparency, fairness, and sector-specific standards designed to protect the processing of data by AI systems and for AI training purposes [136]. More specifically, China has recently developed guidelines for managing AI user privacy [5], supplementing its *Personal Information Protection Law*, which requires explicit consent for data processing and mandates data localization [18]. The centrality of consent in data processing also emerges in Indian legislation, which, through the *Digital Personal Data Protection Act*, strengthens accountability measures for personal data within the context of AI. In India, the Supreme Court's landmark case *K.S. Puttaswamy v. Union of India* further entrenched privacy as a fundamental right, setting constitutional limits on data use and informing safeguards such as data minimization, purpose limitation, and consent in AI applications [137].

In other Asian countries, the approach to privacy protection remains largely focused on soft law. South Korea's *AI Basic Act* and *AI Framework Act* introduce binding obligations for "high-impact" AI systems in sensitive sectors, requiring human-in-the-loop oversight, explainability, and protection of user rights, including privacy [115]. Indonesia's Kominfo Circular No. 9 of 2023 similarly identifies personal data protection as a guiding principle for AI use, though its voluntary nature limits enforceability [9]. Japan has issued a series of non-binding AI guidelines centred on the human-centric development of new technologies, emphasizing the fair and lawful processing of users'

personal data. Vietnam's Decision No. 1290/QD-BKHCN (2024) requires developers to ensure AI systems respect privacy and dignity by preventing discrimination and unfairness in data use [109]. Singapore has taken a different approach with *AI Verify*, a testing and governance toolkit to validate compliance with principles including transparency, fairness, accountability, and data governance [91]. Together, these examples illustrate how Asian states are experimenting with a mix of constitutional rulings, statutory provisions, and soft-law instruments to address privacy concerns in AI, though uneven enforcement remains a significant challenge.

3.1.4 Comparative Analysis

The preceding overview illustrates how approaches to privacy and data protection in AI diverge significantly across regions. In Europe, privacy and data protection are entrenched as enforceable fundamental rights within a comprehensive and unified regulatory framework. In contrast, Asian states present a more heterogeneous picture, ranging from constitutional recognition and binding statutory instruments to non-binding guidelines, voluntary initiatives, and sector-specific toolkits. This diversity reflects deeper differences in legal traditions, governance priorities, and the balance struck between individual rights, state interests, and economic development. A comparative perspective therefore highlights not only the different regulatory logics shaping AI governance in ASEM countries, but also the common technical and institutional challenges that persist across contexts.

Table 2: Comparative Analysis of Privacy and Data Protection in AI

Dimension	Europe	Asia
Regulatory Maturity	Many regulatory binding laws and treaties, e.g. GDPR, AI Act, DSA	Mix of binding laws — as in China or India — and voluntary guidelines, even if soft law remedies are still predominant.
Enforcement and Oversight	Strong supervisory frameworks, with the possibility to export some of the prevision — as in the case of the GDPR.	In many Asian countries supervisory measures are still pretty limited and the exportation of rules is challenged by strategic divergences. Some alignments are still possible, especially thanks o ASEAN guidelines.
Orientation	Human-centric, with an intervention approach predominantly risk-based and focused on impact assessment strategies.	Innovation-centric, with an intervention based on state-guided balance between security and control.

In sum, while Europe advances through a harmonised and enforceable framework that emphasises individual autonomy and consent, Asia demonstrates a spectrum of

responses that combine binding legal measures with soft-law and experimental approaches. Asian approaches often reflect a hybrid model, balancing state and economic interests with privacy protections, contrasting with the EU's rights-based GDPR model. This heterogeneity shows how privacy in AI is framed not only as an individual right but also in relation to collective values, state authority, and developmental priorities. Despite these divergences, both regions face parallel challenges in addressing data-driven risks of AI, including the protection of sensitive information, the enforcement of consent, and the prevention of misuse. The comparative perspective underscores that no single model yet offers a complete solution, and that cross-regional dialogue within ASEM remains essential to developing effective and rights-respecting governance of AI.

3.1.5 General Recommendations

Despite the considerable attention that researchers and technical experts have devoted for years to developing AI systems that respect the fundamental human right to privacy and comply with the imperative to protect users' data, numerous challenges persist. In this regard, some recommendations are outlined below for ASEM countries, aimed at overcoming the risk of mere formal compliance and moving toward the substantive guarantee of these rights:

- **Strengthen enforcement and oversight bodies:** It could be relevant to reinforce the capacity for audits and breach reporting, taking inspirations from the GDPR authorities.
- **Privacy Preserving Technologies:** Directing the attention of technical experts and policymakers toward fostering the development and dissemination of technologies that incorporate mechanisms for protecting and securing users' personal data is paramount. This approach should supplant the prevailing remedial logic that continues to dominate attempts at global data governance. Achieving this objective first requires the widespread promotion of privacy-by-design principles in AI systems.
- **International cooperation:** Favouring international collaboration is of central importance—not only for the creation of multilateral platforms for sharing best practices and facilitating regulatory alignment, but also for the ongoing dialogue concerning regulatory approaches to be adopted. Such cooperation aims to achieve more harmonized legislation concerning privacy protection in the digital era. This is particularly relevant given the coexistence of strict consent-based regimes (e.g., India, China) and more flexible or voluntary models (e.g., Indonesia, Japan), which create mismatches in cross-border data flows.
- **Cross-border enforcement mechanisms:** Strengthening international cooperation is essential not only at the governmental level but also in the development of effective and tangible enforcement tools capable of transcending national boundaries. Since AI services routinely cross jurisdictional borders, reliance on enforcement mechanisms valid only in specific circumstances may encourage forum shopping and undermine certainty regarding the consequences of privacy

infringements or data breaches. Similarly, access to affordable legal remedies is crucial for upholding privacy rights in the global deployment of AI. ASEM dialogue should seek to build bridges between binding and non-binding approaches, exploring how voluntary frameworks and toolkits can complement legal instruments and improve enforcement.

- **Continuous reassessment:** Given the rapid technological advancements in AI, it is imperative to maintain a reiterate assessment of privacy governance frameworks to ensure they remain dynamic and adapted to ongoing innovation.

Approaches within Asia remain heterogeneous, ranging from binding statutory instruments in India, South Korea, and Vietnam, to soft-law initiatives such as ASEAN's Guide on AI Governance and Ethics, Japan's non-binding guidelines, and Singapore's AI Verify toolkit. By contrast, European approaches are largely unified under binding and enforceable instruments such as the GDPR, the revised Convention 108, and the AI Act, which together establish a comprehensive framework centred on individual autonomy, consent, and accountability. This divergence underscores the value of ASEM dialogue in bridging approaches and promoting mutual learning across regions.

3.2 Equality and Non-Discrimination

Equality and non-discrimination are core principles of international human rights law, enshrined in Article 2 of the Universal Declaration of Human Rights (UDHR), Article 26 of the International Covenant on Civil and Political Rights (ICCPR), and Article 21 of the Charter of Fundamental Rights of the European Union (EU Charter). Unlike technical definitions of "fairness" in computer science, international human rights law treats discrimination as an absolute prohibition: there can be no "tolerable" threshold of discriminatory treatment [150, 177].

In the context of AI, the risk that algorithmic systems may reinforce or even exacerbate existing inequalities has become one of the most urgent concerns for policymakers, legal agencies, and civil society alike. There are many sources for bias in AI systems (see section 3.2.1), but data is often seen as the primary medium through which AI systems are trained and thus the fundamental basis for algorithmic predictions or decisions. This signifies that the quality of data exerts a direct influence on the functionality of these systems, independently of other sources of bias. A salient issue is the recognition that data is inherently subject to biases that are intrinsic to it and to those that are context-specific to its generation or utilization [30, 140]. Importantly, intersectional bias, where gender, race, disability, class and other identities compound to produce unique harms, remains a pressing but under-detected problem. These forms of bias directly undermine the right to equality and the essence of human dignity [55, 34]. Consequently, contemporary AI technology may amplify these biases, thereby jeopardizing the fundamental right to equality and non-discrimination.

As we will see in section 3.2.2, across the ASEM region, real-world cases illustrate the breadth of these risks. In Europe, welfare fraud detection tools (SyRI, Netherlands) and grading algorithms (UK Ofqual) have been struck down for systemic discrimination and opacity. In France and Germany, predictive policing and employment profiling

have revealed indirect discrimination and lack of transparency. In Asia, large-scale infrastructures such as Aadhaar in India, credit scoring in China and the Philippines, and AI-driven recruitment in South Korea and Indonesia show how exclusionary outcomes disproportionately affect vulnerable groups, often reframed as technical failures rather than rights violations. But concerns about algorithmic bias are not unique to the Asia–Europe context. Well-documented global cases such as the COMPAS recidivism tool in the United States, which disproportionately misclassified Black defendants as high risk, and Amazon's experimental recruitment algorithm, which systematically disadvantaged female applicants, have become reference points in the international debate. While these cases lie outside the ASEM region, they illustrate the broader mechanisms through which AI systems can replicate and intensify existing inequalities. The following discussion turns to examples from ASEM partner countries, where similar risks have manifested in welfare, policing, education, and employment contexts.

The following subsections examine in greater detail the sources of bias in AI (3.2.1), real-world examples across the ASEM region (3.2.2), legal and policy responses (3.2.3), and comparative insights from Europe and Asia (3.2.4). Building on this analysis, section 3.2.5 identifies general recommendations for ensuring that AI governance fully upholds the non-derogable right to equality and non-discrimination. This chapter ends with a list of questions to be discussed in the Working Group session (3.2).

3.2.1 Sources of Bias in AI Systems

AI systems frequently reproduce or intensify existing social inequalities. Bias enters at different stages, including the composition of training datasets, the design of algorithms, and the contexts in which they are deployed. As a result, groups that are already marginalised face disproportionate harms in welfare, education, policing, and employment. The literature highlights five recurring sources of bias: data bias, algorithmic bias, design bias, deployment bias, and intersectional bias.

Sources of Privacy Harms in AI Systems

- **Data bias**: Training datasets reflect historical inequalities and underrepresentation [17].
- **Algorithmic bias**: Model architectures and optimization amplify unequal outcomes [12].
- **Design process bias**: Lack of diversity in development leads to blind spots in impacts [150].
- **Deployment bias**: Context of use (e.g., policing, credit) can create discrimination even with accurate models [42].
- **Intersectional bias**: Overlapping forms of discrimination across race, gender, or class compound harms [49].

The persistence of these different forms of bias shows that algorithmic discrimination is not an isolated error but a systemic risk. Addressing it requires preventive design, robust monitoring, and effective legal safeguards across sectors.

3.2.2 Illustrative Examples

Across the ASEM region, several high-profile cases have demonstrated how algorithmic decision-making can entrench or magnify structural inequalities. In Europe, the Netherlands' child benefits scandal and the SyRI welfare fraud detection system revealed how opaque profiling in public administration can lead to systemic discrimination, lack of accountability, and severe social harm [131, 14]. In the United Kingdom, the Ofqual grading algorithm disproportionately downgraded students from disadvantaged schools during the COVID-19 pandemic, sparking public outcry and concerns about systemic bias and lack of transparency [159]. In France, predictive policing initiatives such as PAVED raised concerns about opacity and data-driven feedback loops that reinforce discriminatory policing [114]. In Germany, the Federal Employment Agency's "Arbeitsmarktchancen-Index" was criticised for profiling job seekers based on sensitive variables such as age, health, and migration background, thereby risking indirect discrimination and unequal access to social rights [74].

In Asia, India's Aadhaar biometric identification system has been associated with the exclusion of vulnerable groups such as the rural poor, women, and the elderly from essential entitlements, illustrating how large-scale digital infrastructures can reinforce structural inequalities [138]. China's pilot projects for social credit and credit-scoring systems have raised concerns of indirect discrimination, particularly where proxies such as geographic location or social networks are used [89]. In the Philippines, AI-driven credit scoring models risk excluding low-income groups by relying on Western-centric datasets that fail to capture local demographic and linguistic realities [104]. In South Korea, AI-based hiring tools and related systems have faced criticism for opacity, discriminatory outcomes, and biased behaviour in conversational agents such as the chatbot Lee Luda [47]. Similarly, in Indonesia, AI-driven job-matching platforms have been found to disadvantage female applicants due to systemic occupational segregation reflected in training data [132].

Taken together, these cases demonstrate that algorithmic discrimination is not incidental but a systemic risk across welfare, policing, education, employment, and financial services. The examples highlight how historical data, proxy variables, and feedback loops produce errors that are unequally distributed and difficult to correct at scale. Algorithmic bias thus directly intersects with regional human rights protections, making it a pressing governance challenge. For further illustrative cases and detailed sources, see Appendix C.2.

3.2.3 Legal and Policy Responses

Several international and regional legal and policy initiatives have been developed to address the risks of inequality and discrimination in AI. These initiatives generally intertwine legal obligations with soft-law approaches at the intergovernmental level, with the objective of ensuring that AI systems are developed and deployed in a manner that

respects fundamental rights and maintains legal standards of equality. These initiatives converge on some crucial points, such as the attempt to prevent algorithmic discrimination through a mandatory impact assessment on the principle of equality, the promotion of transparency, the adoption of an inclusive approach that favours multi-stakeholder engagement, and the attempt to promote the logic of harms prevention rather then focusing only on compensation or mitigation of damages [68].

At the international level, the UN Office of the High Commissioner for Human Rights (OHCHR) has stressed that biometric surveillance and predictive policing practices raise particular risks of racial and ethnic discrimination [163]. These normative instruments underscore the need for AI governance that directly confronts discrimination, rather than treating it as a secondary risk. Thus, the OHCHR has advocated for thematic investigations and calls for action in 2024, with the objective of promoting non-discrimination throughout the life cycle of AI systems. Specifically, it has endeavoured to encourage the participation of civil society and members of minority groups and those often regarded as marginalized in research concerning the design and development of new intelligent technologies [10]. Furthermore, it has highlighted the accomplishments and shortcomings of the measures implemented thus far to limit or eradicate racism, homophobia, and other forms of intolerance perpetrated by and through AI.

On addressing gender and intersectional bias, instruments such as the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW) and regional equality frameworks provide a legal basis for addressing gendered and intersectional harms [162]. Yet most AI governance instruments to date have not operationalised these obligations, leaving intersectional discrimination insufficiently addressed.

Moreover, the UN´s Global Digital Compact has promoted the establishment of an Independent International Scientific Panel on A and a Global Dialogue on AI Governance which should have the aim of guaranteeing the integration of human rights insights and concrete actors into policy-making at international level, so as to promote a more cohesive AI regulation and respect for the right to equality [126].

Following a similar line, the OECD has developed the so-called OECD AI Principles, the first official attempt at intergovernmental standards on AI, which bind signatories to develop technologies that respect the rule of law, fundamental rights, and democratic values [119]. Among these principles, inclusiveness and respect for and promotion of diversity have been repeatedly emphasised [174]. The European Union appears among the adherents, alongside 44 other countries, including Asian countries such as Singapore, Japan, Korea, and Saudi Arabia.

Among other global initiatives it is also relevant to highlight the Ethically Aligned Design initiative promoted by the Institute of Electrical and Electronics Engineers (IEEE) [21]. It developed ethical guidelines which should guide the design of new technology towards a de-biased and non-discriminatory result. Similarly, the Global Partnership on AI (GPAI) — an international initiative involving over 20 countries, including European ones — promoted research and development of inclusive AI systems, underlining the relevance of favouring non-discrimination in AI governance.

Furthermore, from the point of view of black letter law, many regulations have been introduced at regional level to protect these rights. Recent regulatory developments in Europe are aimed at addressing some of these risks. The EU AI Act includes among

the AI systems which expose to "unacceptable risk" those that enable social scoring or perpetuate discriminatory practices, considering them as prohibited technologies [129]. High-risk systems require a FRIA before deployment [166]. Similarly, the Council of Europe's 2024 Framework Convention on AI and Human Rights obliges signatories to embed safeguards against algorithmic bias, with explicit attention to the protection of vulnerable groups. Article 22 of the GDPR, then, aims to prevent algorithmic discrimination, particularly for decisions that could impact individuals based on characteristics related to gender, religion, ethnicity, or sexual orientation.

Asian countries are experiencing a dynamic interplay between soft law initiatives and more binding statues, even though the option of a uniform regulatory system does not yet appear to be the prevailing approach.

ASEAN has established a set of regional standards that encourage its member states to adopt or enhance legal and policy frameworks: the Guide on AI Governance and Ethics. These frameworks, aim to prevent AI-induced discrimination and promote inclusiveness, particularly in the domains of data governance and algorithmic fairness [136]. Similarly, the South Asian Association for Regional Cooperation (SAARC) set regulatory guidelines to foster equality and prevent the reinforcement of social, racial, and gender divisions [63].

Individual states have also established more specific regulatory and governance plans, which are more preventive than remedial or punitive in nature. Japan, for example, has created advisory institutions with the power to intervene in cases where discriminatory effects or dangers arise from the use of AI systems [79, 38]. Singapore has developed a framework that specifically addresses AI technologies involved in personnel selection processes, guaranteeing recourse for discrimination related to the operation of automated systems and providing fair employment practice guidelines that prohibit all forms of discrimination based on gender, health conditions, age, religion, or marital status [73, 157]. In addition, China has introduced a series of binding regulations since 2021 — notably the Internet Information Service Algorithm Recommendation Management Regulations and the Interim Measures for the Administration of General AI Services — which require data training and data services systems, as well as algorithm design, to develop and implement tools to prevent discrimination [142, 107]. Among these measures, bias audits and ethical impact assessments are primarily promoted [107].

Further recent initiatives reinforce this trend. South Korea's *AI Basic Act* and *AI Framework Act* designate "high-impact" systems in sectors such as healthcare and education, requiring explainability, human oversight, and fairness protections [115]. Vietnam's Decision No. 1290/QD-BKHCN obliges developers to prevent discrimination and unfairness in training datasets, embedding equality safeguards into system design [109]. In India, the *Responsible AI for All* framework emphasizes inclusivity and non-discrimination, drawing on constitutional guarantees of equality and fundamental rights [117]. Taken together, these initiatives show how Asian states are beginning to integrate equality principles into AI governance through both binding legislation and soft-law instruments, though enforcement remains uneven across jurisdictions.

3.2.4 Comparative Analysis

Table 3 provides a comparative analysis of how bias and discrimination in AI systems manifest differently across Europe and Asia, along the following aspects: sectors affected, governance framings, institutional responses, and emerging patterns.

While both regions face common technical sources of bias, skewed datasets, feedback loops, and underrepresentation in design processes, their governance responses diverge. In Europe, algorithmic bias is increasingly framed as a direct violation of the absolute human rights prohibition on discrimination, reinforced by binding legal instruments such as the EU AI Act and the Council of Europe's Framework Convention. Whereas the EU frames bias as a binding human rights violation, Asian systems frequently treat it as a risk management or modernization issue That is, in Asia bias often emerges in large-scale state-led infrastructures such as digital identity and credit-scoring systems, where exclusion is treated primarily as a technical or access issue rather than a rights violation. This contrast reveals not only a difference in legal framing but also in institutional pathways for redress: European courts and regulators actively strike down unlawful systems, while many Asian contexts rely more on judicial review or sectoral regulation.

Table 3: Comparative Analysis of Bias and Non-Discrimination in AI

Dimension	Europe	Asia
Sectors where bias appears	Welfare profiling, policing, education, employment	Digital ID and welfare, credit scoring, hiring, education, surveillance
Typical sources of bias	Historical inequalities reflected in datasets; reliance on sensitive variables (e.g. age, ethnicity, socio-economic status); feedback loops in policing and welfare	Large-scale state-led infrastructures using proxies (e.g. geography, networks); data quality gaps; Western-centric models applied in local contexts; linguistic and demographic under-representation
Governance framing	Bias recognised as discrimination prohibited under EU Charter and ECHR; some AI uses labelled "unacceptable risk" under the AI Act; courts active in striking down unlawful systems	Bias often treated as exclusion or access problem, addressed through judicial review or sectoral regulation; even if several countries introduce explicit fairness and anti-discrimination principles lack of comprehensive anti-discrimination frameworks tailored to AI
Institutional responses	Data protection authorities (ICO, CNIL), courts, Council of Europe instruments, EU AI Act compliance tools (e.g. Fundamental Rights Impact Assessments)	Constitutional courts (e.g. India, South Korea), data protection laws (e.g. PIPL in China), and emerging regulatory instruments (e.g. South Korea's *AI Basic Act*, Vietnam's Decision No. 1290, ASEAN's regional guidance), but oversight remains fragmented and ex ante safeguards are uneven
Key patterns	Bias is formally prohibited but enforcement uneven; strong rights-based framing sometimes clashes with technical "fairness" approaches	Bias risks amplified by scale of infrastructures; systemic inequalities reframed as technical malfunctions rather than rights violations

As Table 3 shows, both Europe and Asia face systemic risks of algorithmic discrimination, but their institutional and normative contexts differ. In Europe, discrimination

is framed as an absolute prohibition under binding human rights instruments such as the EU Charter of Fundamental Rights and the European Convention on Human Rights. This framing rejects any notion of "tolerable" bias and has produced regulatory tools like the AI Act and the Council of Europe's Framework Convention, reinforced by active judicial interventions.

In Asia, bias more often emerges in large-scale state-led infrastructures such as digital identity or credit-scoring systems. Here, exclusion and unequal access are frequently treated as technical or administrative problems, addressed through sectoral regulation or judicial review rather than comprehensive anti-discrimination norms. This fragmented and pragmatic orientation makes systemic inequalities more likely to be framed as access issues rather than rights violations.

For ASEM partners, these divergent approaches create both risks and opportunities. Europe provides clear normative anchors but struggles with uneven enforcement. Asia's scale and diversity highlight urgent risks of exclusion while also offering lessons for stress-testing fairness safeguards. The ASEM process is uniquely positioned to bridge these approaches by harmonising perspectives into actionable safeguards, ensuring that AI systems promote equality and inclusion rather than reinforce structural discrimination.

3.2.5 General Recommendations

Despite notable progress in acknowledging and addressing bias in AI, significant challenges persist. To ensure that AI systems uphold the absolute prohibition of discrimination, ASEM partners should pursue a rights-based strategy that goes beyond technical fixes. The following priority actions are recommended:

- **Institutional safeguards:** Embedding equality impact assessments into procurement, funding, and deployment decisions for AI systems.
- **Inclusive design:** Ensuring meaningful participation of affected groups, including women, minorities, and persons with disabilities, throughout the AI lifecycle.
- **Oversight and accountability:** Strengthening the role of courts, regulators, and independent oversight bodies in scrutinizing discriminatory AI applications.
- **Cross-regional cooperation:** Harmonizing standards through Asia-Europe collaboration, enabling the sharing of methodologies for non-discrimination audits and joint capacity-building initiatives.
- **Public empowerment:** Enhancing AI literacy to enable individuals and communities to understand, contest, and influence how algorithmic systems affect their rights.
- **Intersectional safeguards:** Develop standards and audits that explicitly detect and mitigate overlapping forms of discrimination, such as those based on gender, race, disability, or class, ensuring that AI systems do not reinforce compounded inequalities.

Given the heterogeneity of approaches across Asia, with binding instruments emerging in South Korea, Vietnam, and India, alongside predominantly soft-law frameworks in ASEAN, Japan, and Singapore, future efforts should prioritise strengthening enforcement and oversight capacity to ensure that fairness principles are effectively applied in practice. Regional initiatives such as ASEAN's *Guide on AI Governance and Ethics* (2025) provide an important platform for convergence, but require complementary mechanisms to translate guidance into accountability. In addition, while preventive measures such as bias audits and fairness requirements are increasingly promoted, these need to be accompanied by monitoring and sanctioning tools to ensure that safeguards are not only designed but also enforced.

Asia and Europe, through the ASEM framework, are uniquely positioned to collaborate on joint standards for algorithmic fairness, exchange best practices for non-discrimination audits, and build institutional capacity to monitor compliance. Such cooperation would not only strengthen protection within each region, but also contribute to shaping global norms in rights-based AI governance.

Ultimately, equality and non-discrimination must be treated as non-derogable principles. Unlike risk-based approaches that accept trade-offs, discrimination cannot be tolerated in any form. By embedding this absolute standard into technical design, governance frameworks, and institutional practices, Asia and Europe can lead the way in ensuring AI becomes a tool for inclusion rather than exclusion, demonstrating global leadership in developing technologies that serve all members of society fairly and justly.

3.3 Remedies and Access to Justice

The right to an effective remedy is a cornerstone of international human rights law. It is affirmed in Article 8 of the Universal Declaration of Human Rights (UDHR), Article 2(3) of the International Covenant on Civil and Political Rights (ICCPR), and regional instruments such as Article 47 of the Charter of Fundamental Rights of the European Union. These provisions establish that where rights are violated, individuals must have timely, accessible, and enforceable avenues of redress. As outlined in Chapter 2, both international and regional frameworks recognise remedies as essential for ensuring that rights protections are not merely declaratory but practically enforceable.

In the AI context, this principle faces distinctive challenges. Algorithmic harms often arise from opaque, complex, and multi-actor systems that make it difficult to identify responsibility or contest outcomes [173, 41]. Unlike traditional rights violations, AI-related harms are frequently diffuse (affecting groups rather than individuals), transnational (spanning multiple jurisdictions), and systemic (embedded in infrastructures and processes rather than isolated acts) [46]. These features strain existing legal and institutional mechanisms of redress, and raise critical questions about whether existing human rights frameworks are sufficient or whether new legal tools are required [103].

The following subsection (3.3.1) identifies the main barriers that obstruct effective remedies in practice, ranging from opacity and diffusion of responsibility to collective harms, jurisdictional complexity, and resource gaps.

3.3.1 Barriers to Remedies

When AI causes harm, access to justice is often obstructed by structural and procedural barriers. These challenges are not incidental but stem from the technological, legal, and socio-economic characteristics of AI systems [166, 41]. Opacity in complex models makes it difficult to detect harms or contest outcomes, with some arguing that the absence of explainability should itself be treated as a rights violation [168]. Responsibility is frequently diffused across developers, deployers, regulators, and data providers, creating accountability gaps where no single actor can be held liable [16, 20]. Existing frameworks are also poorly suited to collective and systemic harms, such as predictive policing or biased grading, because they are largely designed for individual claims [98]. The transnational nature of AI complicates accountability by raising jurisdictional conflicts and uncertainty about applicable law. Finally, power asymmetries leave affected communities with limited resources to pursue justice, while novel harms like predictive profiling, reputational damage, or anticipatory surveillance often fall outside established legal categories.

Barriers to Effective Remedies in AI

- **Opacity:** black-box systems prevent individuals from contesting harms [168, 173].
- **Diffuse responsibility:** unclear accountability across many actors [41, 103, 36].
- **Collective and systemic harms:** group-level discrimination lacks clear remedies [43].
- **Jurisdictional complexity:** cross-border AI use blurs applicable law [103, 175].
- **Power asymmetries:** individuals lack resources to challenge powerful deployers [41, 106].
- **Novel harms and legal mismatch:** harms that fall outside traditional legal categories, can leave victims without clear redress [46, 135, 168].

These barriers underscore why remedies must evolve beyond compensation toward procedural safeguards, accountability measures, collective redress, and structural reforms that guarantee timely and effective justice [103].

3.3.2 Illustrative Examples

Examples from Asia and Europe highlight the urgent need for effective remedies. Some of the same cases discussed in Section 3.2.2 are revisited here, not to re-examine harms, but to show the remedial pathways pursued in practice, such as judicial annulments, disclosure rights, and regulatory enforcement. For further illustrative cases and detailed sources, see Appendix C.3.

Beyond the ASEM region, courts and regulators have also acted decisively. In Kenya, the High Court suspended the Huduma Namba biometric ID system until a proper data protection framework, including a DPIA, was in place (2021) [76]. In Canada, privacy commissioners required Clearview AI to stop collecting facial recognition data and delete existing databases (2021), followed by an Alberta court ordering the cessation of services and deletion of data in that province (2025) [123, 51]. These actions illustrate how preventive and enforcement remedies can operate both structurally and ex post.

Taken together, the cases show that remedies for AI-related harms are beginning to take shape across diverse legal systems. Courts have annulled unlawful systems, mandated disclosure, and embedded procedural safeguards, while regulators have imposed audits, transparency obligations, and structural reforms. Yet remedies remain uneven, often reactive, and dependent on litigation or regulatory discretion. For ASEM partners, the challenge is to move from ad hoc responses to coherent, proactive frameworks that ensure timely, accessible, and effective remedies for both individual and collective harms.

3.3.3 Emerging Mechanisms

Several legal and institutional mechanisms are beginning to address the need for remedies in AI-related harms. Their development is uneven across regions, but some common patterns are visible.

Europe. Existing frameworks such as the GDPR already provide enforceable rights to information, access, correction, deletion, and complaint to data protection authorities. Courts have enforced these rights in cases such as *SyRI* in the Netherlands and administrative disclosure rulings in France. More recently, the EU AI Act (2024) introduces ex ante safeguards, including mandatory Fundamental Rights Impact Assessments (FRIA) for high-risk systems. National data protection authorities, ombuds institutions, and national human rights institutions are gradually expanding their mandates to include algorithmic grievances. The Council of Europe's Framework Convention (2024) embeds binding obligations to provide remedies for rights violations linked to AI.

Asia. Emerging mechanisms are more fragmented. Some jurisdictions rely on binding laws: China's Personal Information Protection Law (2021) allows individuals to seek civil remedies for unlawful data use, and India's Digital Personal Data Protection Act provides access, correction, and erasure rights. Japan and South Korea employ a mix of soft law guidelines and constitutional protections, with courts increasingly scrutinising surveillance and data use. Regulatory authorities, such as the Philippines' National Privacy Commission, have begun imposing fairness audits or corrective measures in financial and education sectors. However, collective remedies remain rare, and access is often limited to administrative or sectoral channels. In the Philippines, Bill No. 7396 (2024) proposes the creation of the Artificial Intelligence Development Authority (AIDA), which would regulate AI and provide accessible complaint mechanisms for individuals affected by AI-related harms [66]. In South Korea, the *AI Framework Act* introduces incentives for developers to conduct voluntary Human Rights Impact Assessments (HRIAs), linking such practices to eligibility for public procurement

processes [115].

Furthermore, at the international level, the OHCHR has stressed that access to remedies must explicitly cover AI-driven harms, urging states to strengthen institutional mandates to provide both individual and collective redress [163].

Convergence. ASEM regions are experimenting with algorithmic impact assessments, expanding regulator mandates, and strengthening ex ante oversight. Judicial willingness to intervene is also increasing, signalling a slow but notable recognition that effective remedies are essential to safeguarding rights in the AI era.

3.3.4 Comparative Analysis

The comparative analysis of Asia and Europe in Table 4 reveals convergences and divergences in how remedies are conceptualised and implemented.

Convergences. In both regions, opacity is the central barrier, and remedies increasingly focus on transparency obligations and disclosure rights. Regulators play a growing role, while courts are willing to strike down opaque or rights-infringing systems. There is also a gradual shift from purely compensatory remedies toward structural and procedural safeguards, such as impact assessments and audit requirements.

Divergences. Europe frames remedies through the lens of enforceable human rights-effective remedy, fair trial, and non-discrimination. Courts and regulators explicitly treat algorithmic harms as rights violations, enabling systemic remedies and collective redress in some instances. By contrast, Asian approaches often frame remedies as matters of consumer protection, administrative oversight, or technical compliance. Structural or collective harms are less often recognised as rights violations, and remedies are typically limited to individual grievances.

Reflections. Despite progress, remedies remain reactive and uneven. The burden of proof continues to fall on individuals, even when harms are opaque and systemic. Ex ante tools such as FRIAs offer promise but depend on strong enforcement. Participation of affected communities in shaping remedial mechanisms is limited, while cross-border AI services highlight jurisdictional gaps that neither region has adequately addressed. For ASEM partners, the challenge is to move beyond fragmented, ad hoc remedies toward proactive, harmonised frameworks that guarantee timely, accessible, and enforceable redress for both individual and collective harms.

Table 4: Comparative Analysis of Remedies for AI-related Harms

Dimension	Europe	Asia	Global / International
Legal framing	Remedies framed as enforceable human rights (effective remedy, fair trial, non-discrimination). AI Act and Council of Europe Convention embed ex ante safeguards.	Remedies often framed as consumer protection, administrative oversight, or technical compliance. Fragmented laws (e.g. PIPL in China, DPDP in India, soft law in Japan).	OHCHR and UN bodies stress access to remedies as a human rights obligation, calling for both individual and collective redress. No binding global framework yet.
Institutions	Data protection authorities (DPAs), ombuds offices, national human rights institutions, courts. Strong ex ante powers in AI Act.	National DPAs and sectoral regulators (financial, telecom, education). Courts active in surveillance/privacy. NHRI involvement limited.	OHCHR, UN treaty bodies, UNESCO, OECD promote guidance and monitoring; no global enforcement body.
Types of remedies	Individual rights (access, correction, deletion), procedural safeguards (impact assessments, disclosure rights), collective redress in some contexts (class actions, public interest litigation).	Primarily individual remedies: access, erasure, correction. Emerging mechanisms include AI disclosure duties (China), proposed complaint authority (Philippines), and voluntary HRIAs (South Korea), but limited in scope and enforcement	Normative emphasis on both individual and collective remedies; structural remedies recommended (mandating disclosure, institutional strengthening), but remain aspirational.
Common gaps	Enforcement uneven across Member States; remedies often reactive; burden of proof still on individuals; cross-border enforcement weak.	Fragmented, uneven enforcement; limited recognition of systemic harms; heavy reliance on administrative or soft law solutions.	Lack of binding obligations; reliance on state cooperation; limited monitoring and no direct enforcement.
Emerging trends	Mandatory Fundamental Rights Impact Assessments (FRIAs), stronger regulator mandates, courts striking down unlawful AI systems.	Increasing judicial scrutiny (surveillance, welfare exclusion), stronger data protection laws in China and India, regulator-led audits.	Growing recognition of AI harms in UN forums; proposals for global AI governance and remedy standards; emphasis on transnational cooperation.

3.3.5 General Recommendations

In Europe, remedies for AI-related harms are increasingly embedded in binding legal frameworks, with the GDPR, national data protection laws, and the forthcoming AI Act providing enforceable rights and procedural safeguards, complemented by active judicial oversight. While some Asian states have begun to experiment with AI-specific remedies, such as disclosure requirements in litigation (China), proposed complaint mechanisms (Philippines), and incentives for voluntary Human Rights Impact Assessments (South Korea), these initiatives remain limited in scope and enforcement. Strengthening these efforts and ensuring their alignment with international human rights standards should be a priority for ASEM dialogue. Building on the comparative analysis in the previous section, several priority actions emerge for ASEM partners to ensure that remedies for AI-related harms are timely, accessible, and enforceable:

- **Transparency as a prerequisite:** Disclosure and explainability must be recognised as preconditions for access to remedies. The absence of explainability can itself constitute a violation of rights, since it prevents contestation. Tools such as mandatory Fundamental Rights Impact Assessments (FRIAs) and a statutory right to explanation should be embedded in legal frameworks.
- **Clarifying accountability chains and lowering barriers:** Legal frameworks should clearly allocate responsibilities across developers, deployers, and regulators to avoid responsibility gaps. At the same time, burdens of proof and litigation costs must not fall disproportionately on affected individuals, particularly when harms are opaque and systemic.
- **Collective remedies:** Beyond individual claims, systemic harms require mechanisms such as group litigation, public interest actions, and systemic investigations by regulators or national human rights institutions.
- **Participation of affected communities:** Remedies should be designed with meaningful participation of those most affected. Mechanisms such as citizen juries, consultation processes, and design justice frameworks enhance legitimacy and ensure that technologies align with social values.
- **Institutional strengthening and cross-border cooperation:** Ombuds offices, data protection authorities, and national human rights institutions must be resourced and empowered to adjudicate AI grievances. Given the transnational nature of AI, cross-border cooperation among regulators is essential to prevent jurisdictional gaps.
- **AI literacy for justice:** Capacity-building for communities, lawyers, judges, and regulators is needed to ensure meaningful access to remedies. AI literacy programmes can help individuals detect harms, contest outcomes, and engage in systemic oversight.
- **State and global responsibilities:** National governments must guarantee effective remedies for inviolable rights, while international cooperation—guided by

OHCHR and other bodies—is vital to ensure that cross-border algorithmic harms are not left without redress.

Ultimately, access to justice must be treated as a non-derogable right. Remedies cannot be optional or symbolic; they are the condition that makes all other human rights protections meaningful. By embedding transparency, accountability, and collective redress into legal and institutional frameworks, ASEM partners can ensure that rights remain not merely declaratory but practically enforceable in the age of AI.

4 The way forward

The rapid expansion of AI technologies presents both opportunities and risks for human rights across Asia and Europe. To ensure that AI development and deployment serve the public good, a coherent and inclusive governance approach is needed, embedding human rights protections into technical design, legal regulation, and institutional oversight.

4.1 Integrating Human Rights in AI Governance:

A growing body of research highlights the need for AI governance to be grounded in shared ethical principles, supported by dynamic regulatory frameworks, and developed through inclusive, multistakeholder processes [70, 3]. Despite the proliferation of AI ethics guidelines, recent systematic reviews underscore that such guidelines remain fragmented in quality and enforceability [26]. Furthermore, a systematic literature review evaluated 61 AI governance studies and found that only a few comprehensively address who governs what, when, and how, underscoring need for integrated frameworks [13]. This gap between principle and practice is particularly relevant for regions like Asia and Europe, where diverse institutional approaches must converge to address transnational human rights risks.

These insights echo the dimensions framework proposed by Xanthopoulou et al. (2025), which underlines that meaningful AI governance requires attention to four key dimensions: the issuing body, scope, application conditions, and governance approach [172], which help differentiate between binding instruments and soft-law tools The study also reveals how impactful initiatives tend to blend legal enforceability with value-driven, participatory mechanisms.

Concrete tools, including algorithmic impact assessments, transparency standards (e.g. the UK's Algorithmic Transparency Recording Standard), and audit frameworks, are crucial for translating abstract commitments into actionable safeguards. These mechanisms, however, must be embedded within institutional structures that guarantee accountability, public oversight, and access to remedies [3].

Crucially, the false dichotomy between innovation and regulation must be rejected. As often argued [146, 3], well-designed governance mechanisms do not inhibit innovation but are core to create the trust and legitimacy necessary for sustainable adoption of AI technologies. As Virginia Dignum has argued, "regulation is innovation", i.e. not an option, rbut a stepping stone that fosters public trust, societal acceptance, and responsible adoption of AI technologies [37].

In both Asia and Europe, multi-stakeholder initiatives are emerging as promising models for embedding human rights in AI governance. In Europe, the High-Level Expert Group on AI and national AI observatories (e.g. in France and Germany) have created structured channels for dialogue between policymakers, academia, industry, and civil society. In Asia, initiatives such as Japan's AI Governance Guidelines and Singapore's Model AI Governance Framework actively involve industry associations and civil society organisations in shaping standards. For ASEM partners, exchanging experiences from these multi-stakeholder processes can help identify best practices for inclusive participation, co-regulation, and the monitoring of human rights impacts across regions.

4.2 Future Directions for AI and Human Rights

The intersection of AI and human rights is evolving rapidly, driven by technological advancement, geopolitical shifts, and new legal frameworks. This section outlines key emerging trends, opportunities, and risks that will shape the future of rights-based AI governance.

Emerging Trends

- **Efforts in AI legislation:** Instruments such as the EU AI Act and the Council of Europe's Framework Convention are setting global benchmarks for rights-based regulation, potentially triggering normative diffusion across regions.
- **Integration of rights-based design:** Increasing incorporation of Fundamental Rights Impact Assessments (FRIAs) into high-risk AI systems reflects a shift toward embedding human rights from the outset.
- **AI for the public interest:** AI is being leveraged for social good in areas such as disaster response, climate monitoring, and public health, with potential to support the realization of economic and social rights.
- **Participatory governance models:** There is growing recognition of the need for inclusive, multistakeholder approaches involving civil society, academia, and marginalized communities.
- **Convergence with environmental and intergenerational concerns:** New governance frameworks increasingly link human rights with sustainability and long-term societal resilience.

Opportunities

- Embedding enforceable human rights protections into the lifecycle of AI systems, including through public procurement and technical standards.
- Enhancing access to justice via AI tools for legal assistance, translation, and information, supported by transparency and human oversight.

- Fostering Asia-Europe leadership in collaborative, rights-based AI governance, informed by shared values and regulatory innovations.
- Strengthening capacity building and knowledge transfer between ASEM partners to support context-sensitive AI governance, especially in emerging economies.
- Establishing global benchmarks and interregional dialogues to promote human-centric AI development.

Risks

- **Opacity and lack of accountability:** Many AI systems remain non-transparent, limiting individuals' ability to understand or contest decisions that affect them.
- **Algorithmic discrimination:** Inadequate representation in data and design processes may reinforce existing inequalities, particularly against marginalized groups.
- **Surveillance overreach:** The unchecked use of AI in biometric identification, predictive policing, and profiling poses serious threats to privacy and civil liberties.
- **Technological dependency:** Reliance on foreign-developed AI systems risks exacerbating digital colonialism and undermining local autonomy.
- **Regulatory fragmentation:** Diverging national and regional approaches may weaken the enforceability and universality of human rights standards in AI governance.

Looking ahead, regional contexts shape both risks and opportunities. In Europe, the development of binding legal frameworks such as the AI Act reflects a rights-based approach anchored in the EU Charter of Fundamental Rights. In Asia, governance approaches are more diverse: some countries (e.g. China, India, Korea) have introduced binding measures, while others rely primarily on soft-law guidelines. This divergence creates opportunities for cross-regional learning, as Europe can share lessons from rights-based regulation, while Asia's experiences with large-scale deployment and rapid innovation highlight the importance of context-sensitive safeguards.

A persistent gap, however, concerns capacity. Several ASEM partner countries, particularly in the Global South, face resource and expertise constraints that hinder the enforcement of remedies or the integration of human rights into AI governance. Addressing this imbalance requires targeted investment in regulatory capacity, judicial training, and technical skills development, alongside mechanisms for peer learning and knowledge exchange across ASEM.

As AI-related harms increasingly cross borders, future governance must integrate not only preventive safeguards but also robust remedial mechanisms. Developing interoperable standards for remedies, building judicial and regulatory capacity, and piloting cross-regional redress mechanisms are key areas where ASEM cooperation can add value.

4.3 Opportunities for Asia-Europe Collaboration

ASEM countries are well-positioned to foster a collaborative, rights-based AI governance model that includes:

- Harmonizing normative standards while accommodating regional and cultural diversity, drawing from instruments such as the EU AI Act and regional Asian frameworks.
- Building institutional capacity for oversight, redress, and enforcement through joint training programmes, regulatory sandboxes, and public-private partnerships.
- Promoting meaningful engagement by civil society, academia, and marginalized groups, with an emphasis on inclusive governance and participatory mechanisms.
- Facilitating inter-regional exchanges on best practices, regulatory innovations, and rights-based tools via ASEM-led platforms, observatories, or annual forums on AI and human rights.
- Developing shared AI audit and assessment frameworks to support mutual accountability and enable cross-border trust in AI systems.
- Encouraging responsible innovation through co-investment in AI for public good projects (e.g. health, disaster response, climate action) that serve both development goals and human rights agendas.
- Supporting AI literacy and human rights education initiatives, especially in low-resource settings, through coordinated efforts in curriculum development, digital training, and knowledge hubs.
- Collaborating on standard-setting in multilateral forums (e.g. UNESCO, OECD, UN bodies) to advance a common ASEM voice on human rights in AI governance.

Such cooperation can help bridge normative, technical, and institutional divides across regions, reinforcing AI systems that are not only innovative but also legitimate, fair, and rights-compliant.

Existing initiatives provide concrete entry points for collaboration. The EU–ASEAN Digital Partnership (2022), the EU–Japan Digital Partnership, and ASEM-wide digital literacy programmes demonstrate that cross-regional cooperation is already underway. These initiatives could be expanded to include explicit human rights benchmarks for AI, joint audit frameworks, and regular Asia–Europe policy dialogues on remedies and access to justice. Moreover, recent Asian frameworks, such as the Chongqing Consensus and China's AI Capacity-Building Action Plan, explicitly frame AI as an international public good and call for cross-regional cooperation [120].

In the future, the following are key areas where Asia and Europe can cooperate to align their efforts, bridge regulatory and developmental gaps, and ensure AI development does not come at the cost of fundamental rights and values:

Aligning Ethical and Governance Frameworks: Asia and Europe share core principles in AI ethics — transparency, accountability, fairness, and human-centricity. A key opportunity lies in aligning these values within interoperable governance frameworks that facilitate innovation while preventing harm. As above mentioned, Europe's binding regulatory models set important legal precedents, while Asia's initiatives offer flexible, context-sensitive approaches. Harmonizing these efforts through dialogue and mutual recognition can strengthen global standards for ethical AI.

Promoting Responsible and Inclusive AI Deployment: Asia and Europe, with their complementary strengths in technology, innovation, and policy-making, have a unique opportunity to collaborate in ensuring that AI serves the common good. By pooling resources and knowledge, the two regions can leverage AI to address critical shared challenges, such as improving healthcare access, reducing educational inequality, enhancing labor rights, and advancing climate resilience.

Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet affirmed six priorities in enabling AI to be human rights based and trustworthy[1]:

Through such cooperation, Europe's strong regulatory frameworks can complement Asia's rapid technological advancement, creating an AI-for-good application ecosystem that is diverse and inclusive.[2]

Fostering Multi-stakeholder Dialogues and Engagement: Multi-stakeholder dialogues can foster knowledge exchange between regions and sectors, strengthening local capacities. By holding joint workshops, conferences, and collaborative research projects, Asia and Europe can share best practices on AI and human rights, such as how to handle algorithmic transparency, data privacy, or the social impact of automation. This capacity building can ensure equitable AI deployment and policy development globally.

Capacity-building and Skills Exchange: Capacity-building should be the central pillar of cooperation. Joint training for businesses, public officer and civil society individuals should continue, but with a deeper focus on more niche areas of AI. For example, academic and civil society exchange programmes aimed at fostering knowledge transfer on algorithmic auditing and rights-based design. One step further would be for ASEM partners to establish a dedicated observatory on AI and human rights to facilitate ongoing exchange of practices, data and methodologies.

5 Conclusions and Recommendations

The analysis in this paper shows that while AI offers significant opportunities for both Asia and Europe, it also poses serious risks to the protection of human rights, particularly in the areas of privacy, non-discrimination, and access to remedies. ASEM

[1]Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet, posted on 11 February 2025 in Paris at the AI Action Summit. https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet.

[2]The 2025 World artificial intelligence (AI) Conference and High-Level Meeting on Global AI Governance published Global AI Governance Action Plan on July 26: https://www.en84.com/16169.html

partners therefore share a responsibility to ensure that AI is developed and deployed in ways that uphold international human rights standards.

In moving forward, it is important that ASEM does not only exchange experiences but also identifies concrete entry points for sustained cooperation. Building on existing EU–ASEAN and EU–Japan partnerships, ASEM could mainstream human rights benchmarks into ongoing digital cooperation frameworks. To operationalise this, several priority actions can be envisaged:

- Establishing an **ASEM Observatory on AI and Human Rights** to facilitate joint monitoring, data-sharing, and policy learning across member states.
- Launching **joint training programmes** for regulators, judges, ombuds institutions, and civil society actors to strengthen capacity in assessing and remedying AI-related harms.
- Piloting **cross-border AI audits or certification schemes** that integrate human rights safeguards, enhancing trust and interoperability of AI governance across regions.
- Supporting **multi-stakeholder dialogue platforms** under ASEM to ensure that voices from academia, industry, and civil society inform governance practices in both Asia and Europe.

These recommendations are intended to provide cross-cutting guidance for ASEM partners, directly addressing the three thematic areas of this paper: privacy and data protection, equality and non-discrimination, and remedies and access to justice; thereby supporting the discussions of the Working Groups.

In pursuing these actions, cooperation should be firmly anchored in existing international human rights obligations, including the ICCPR, the ICESCR, the UN Guiding Principles on Business and Human Rights, and the UNESCO Recommendation on the Ethics of Artificial Intelligence. This ensures that joint initiatives under ASEM reinforce, rather than duplicate, globally recognised standards.

Finally, capacity building should not only focus on institutional actors such as regulators, judges, and ombuds institutions, but also extend to affected communities and marginalised groups. Ensuring their meaningful participation in AI governance will help ASEM partners to design remedies and safeguards that are both inclusive and effective.

Such initiatives would enable ASEM partners to translate high-level commitments into practical cooperation, reinforcing their shared responsibility to ensure that AI serves as a driver of human rights protection and sustainable development.

In sum, the rapid spread of AI across Asia and Europe makes it imperative for ASEM partners to act jointly in embedding human rights into governance frameworks. By addressing risks to privacy, equality, and access to justice in a coherent and coordinated way, and by anchoring cooperation in international human rights standards, ASEM can ensure that AI development strengthens, rather than undermines, democratic values and human dignity. The concrete steps outlined above — from observatories and training programmes to cross-border audits and inclusive dialogue platforms

— provide an actionable path forward. Taken together, these initiatives offer ASEM the opportunity to demonstrate global leadership in aligning technological innovation with the protection and promotion of fundamental rights.

6 Acknowledgement

The contents of this document are the sole responsibility of the authors and can under no circumstances be regarded as reflecting the views or opinions of the organisers or co-funders of the 23rd Informal ASEM Seminar on Human Rights, the Asia-Europe Foundation (ASEF), the Raoul Wallenberg Institute, the Philippine Department of Foreign Affairs, the Swiss Federal Department of Foreign Affairs, the Ministry of Foreign Affairs of the People's Republic of China, and the Ministry of Foreign Affairs of Denmark.

This document has been produced with the financial assistance of the co-organisers and the European Union.

References

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology*, 30(4):736–758, 2020.

[2] Tareq Al-Billeh, Ruba Hmaidan, Ali Al-Hammouri, and Mohammed AL Makhmari. The risks of using artificial intelligence on privacy and human rights: Unifying global standards. *Jurnal Media Hukum*, 31(2):333–350, 2024.

[3] Pekka Ala-Pietilä and Nathalie A Smuha. A framework for global cooperation on artificial intelligence and its governance. In *Reflections on artificial intelligence for humanity*, pages 237–265. Springer, 2021.

[4] Abdulmajeed Alahmari and Bob Duncan. Cybersecurity risk management in small and medium-sized enterprises: A systematic review of recent evidence. In *2020 international conference on cyber situational awareness, data analytics and assessment (CyberSA)*, pages 1–5. IEEE, 2020.

[5] Francisca Romana Nanik Alfiani and Faisal Santiago. A comparative analysis of artificial intelligence regulatory law in asia, europe, and america. In *SHS Web of Conferences*, volume 204, page 07006. EDP Sciences, 2024.

[6] Marco Almada and Nicolas Petit. The eu ai act: a medley of product safety and fundamental rights? *Robert Schuman Centre for Advanced Studies Research Paper*, (2023/59), 2023.

[7] Fionnuala Ni Aolain. The rise of counter-terrorism and the demise of human rights. *Emory Int'l L. Rev.*, 39:1, 2024.

[8] Naomi Appelman, Ronan Ó Fathaigh, and Joris van Hoboken. Social welfare, risk profiling and fundamental rights: The case of syri in the netherlands. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law (JIPITEC)*, 12:257, 2021.

[9] Joshua Ramoti Ariesto. Ai ethics in indonesia: Should ai behave ethically like humans, January 2024. ARFP Law Firm Blog.

[10] KP Ashwini. Contemporary forms of racism, racial discrimination, xenophobia and related intolerance: Report of the special rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, ashwini kp. 2024.

[11] United Nations. General Assembly. *Universal declaration of human rights*, volume 3381. Department of State, United States of America, 1949.

[12] Solon Barocas and Andrew D Selbst. Big data's disparate impact. *Calif. L. Rev.*, 104:671, 2016.

[13] Amna Batool, Didar Zowghi, and Muneera Bano. Responsible ai governance: A systematic literature review. *ArXiv preprint*, 2023.

[14] Sonja Bekker. Fundamental rights in digital welfare states: The case of syri. *Netherlands Yearbook of International Law 2019: Yearbooks in International Law: History, Function and Future*, 50:289, 2020.

[15] Lucas Bergkamp. Eu data protection policy: the privacy fallacy: adverse effects of europe's data protection policy in an information-driven economy. *Computer Law & Security Review*, 18(1):31–47, 2002.

[16] Reuben Binns. Data protection impact assessments: a meta-regulatory approach. *International Data Privacy Law*, 7(1):22–35, 2017.

[17] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81:1–15, 2018.

[18] Igor Calzada. Citizens' data privacy in china: The state of the art of the personal information protection law (pipl). *Smart Cities*, 5(3):1129–1150, 2022.

[19] Celso Cancela-Outeda. The eu's ai act: A framework for collaborative governance. *Internet of Things*, 27:101291, 2024.

[20] Corinne Cath. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180080, 2018.

[21] Raja Chatila, Kay Firth-Butterfield, and John C Havens. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems version 2. 2018.

[22] Chen Junji. Toward agile governance: Exploring japan's artificial intelligence governance model[j], 2024.

[23] Digi China. Internet information service algorithmic recommendation management provisions (2021). 2021.

[24] European Commission. Proposal for a council decision authorising member states to ratify, in the interest of the european union, the protocol amending the council of europe convention for the protection of individuals with regard to automatic processing of personal data (ets no. 108), com(2018) 451 final, 5 June 2018.

[25] European Commission. Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions. ai continent action plan, com(2025) 165 final, 9 April 2025.

[26] Nicholas Kluge Corrêa, Camila Galvão, James William Santos, Carolina Del Pino, Edson Pontes Pinto, Camila Barbosa, Diogo Massmann, Rodrigo Mambrini, Luiza Galvão, Edmund Terem, et al. Worldwide ai ethics: A review of 200 guidelines and recommendations for ai governance. *Patterns*, 4(10), 2023.

[27] Bart Custers and Daniel Bachlechner. Advancing the eu data economy: Conditions for realizing the full potential of data reuse. *Information Polity*, 22(4):291–309, 2017.

[28] Bart Custers and Gianclaudio Malgieri. Priceless data: Why the eu fundamental right to data protection is at odds with trade in personal data. *Computer Law & Security Review*, 45:105683, 2022.

[29] Brian Daigle and Mahnaz Khan. *The changing tides of data protection regulation and enforcement in Europe*. Office of Industries, US International Trade Commission, 2022.

[30] Tushar Kanti Das and Bing-Sheng Teng. Cognitive biases and strategic decision processes: An integrative perspective. *Journal of management studies*, 36(6):757–778, 1999.

[31] Paul De Hert and Guillermo Lazcoz. When GDPR-principles blind each other: accountability, not transparency, at the heart of algorithmic governance. *Eur. Data Prot. L. Rev.*, 8:31, 2022.

[32] Katerina Demetzou. Data protection impact assessment: A tool for accountability and the unclarified concept of 'high risk'in the general data protection regulation. *Computer Law & Security Review*, 35(6):105342, 2019.

[33] Sonia Desmoulin-Canselier and Daniel Le Métayer. Algorithmic decision systems in the health and justice sectors: certification and explanations for algorithms in european and french law. *European Journal of Law and Technology*, 9(3), 2018.

[34] Hannah Devinney, Jenny Björklund, and Henrik Björklund. We don't talk about that: case studies on intersectional analysis of social bias in large language models. In *Workshop on Gender Bias in Natural Language Processing (GeBNLP), Bangkok, Thailand, 16th August, 2024.*, pages 33–44, 2024.

[35] Assunta Di Martino. Robotica medica. *Amministrativ@ mente-Rivista di ateneo dell'Università degli Studi di Roma "Foro Italico"*, (3-4), 2017.

[36] Assunta Di Martino et al. *Intelligenza artificiale e responsabilità civile in ambito sanitario*. Giuffrè Francis Lefebvre, 2022.

[37] Virginia Dignum. Beyond the ai race: why global governance is the greatest innovation. *AI Policy Exchange Forum (AIPEX)*, 2025.

[38] Nicole Dirksen and S Takahashi. Artificial intelligence in japan 2020. *Actors, Market, Opportunities and Digital Solutions in a Newly Transformed World. Netherlands Enterprise Agency*, 2020.

[39] Pam Dixon. A failure to "do no harm"–india's aadhaar biometric id program and its inability to protect privacy in relation to measures in europe and the us. *Health and technology*, 7(4):539–567, 2017.

[40] Dominic Paulger. Understanding japan's ai promotion act: An 'innovation-first' blueprint for ai regulation, 5 July 2025.

[41] Lilian Edwards and Michael Veale. Slave to the algorithm? why a'right to an explanation'is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, 16:18, 2017.

[42] Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Runaway feedback loops in predictive policing. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 160–171. ACM, 2018.

[43] Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

[44] Elaine Fahey. Data protection and regulation of social media. In *Handbook of European Union Governance*, pages 143–155. Edward Elgar Publishing, 2025.

[45] Yang Feng. The future of china's personal data protection law: challenges and prospects. *Asia Pacific Law Review*, 27(1):62–82, 2019.

[46] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, et al. Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Minds and machines*, 28(4):689–707, 2018.

[47] Association for Progressive Communications. Controversial cases of ai in the republic of korea. https://www.apc.org/en/pubs/controversial-cases-ai-republic-korea, 2021. Accessed: 2025-08-20.

[48] Catherine Forget. A challenge to systematic and undifferentiated data collection through strategic litigation: The passenger name record case (ligue des droits humains) before the court of justice of the eu. *German Law Journal*, 25(6):1068–1079, 2024.

[49] James R Foulds, Rashidul Islam, Kamrun Naher Keya, and Shimei Pan. An intersectional definition of fairness. In *2020 IEEE 36th international conference on data engineering (ICDE)*, pages 1918–1921. IEEE, 2020.

[50] A Michael Froomkin. Big data: Destroyer of informed consent. *Yale JL & Tech.*, 21:27, 2019.

[51] Norton Rose Fulbright. Alberta court orders deletion of clearview ai data and cessation of services. *Legal Alert*, 2025. Finds Clearview's use of FRT in Alberta unlawful; mandates data deletion.

[52] Gloria González Fuster. *The emergence of personal data protection as a fundamental right of the EU*, volume 16. Springer Science & Business, 2014.

[53] Andrea Gadotti, Luc Rocher, Florimond Houssiau, Ana-Maria Creţu, and Yves-Alexandre De Montjoye. Anonymization: The imperfect science of using data while preserving privacy. *Science advances*, 10(29):eadn7053, 2024.

[54] Raphaël Gellert. *The risk-based approach to data protection*. Oxford University Press, 2020.

[55] Usman Gohar and Lu Cheng. A survey on intersectional fairness in machine learning: Notions, mitigation, and challenges. *arXiv preprint arXiv:2305.06969*, 2023.

[56] Governament of China. Provisions on the administration of security vulnerabilities in network products, 2021.

[57] Governament of China. Interim measures for the administration of generative artificial intelligence services, 2022.

[58] Governament of China. Provisions on the administration of deep synthesis in internet information services, 2022.

[59] Government of China. Human rights action plan of china (2021-2025), 2021.

[60] Graham Greenleaf. *Asian data privacy laws: trade & human rights perspectives*. Oup Oxford, 2014.

[61] Giuseppe Grosso. *Corso di diritto romano: Le cose*. Giappichelli Torino, 1941.

[62] António Guterres. Roadmap for digital cooperation. *United Nations*, 2020.

[63] Mahmud Hasan. Regulating artificial intelligence: A study in the comparison between south asia and other countries. *Legal Issues in the digital Age*, (1):122–149, 2024.

[64] Louis Henkin. The universality of the concept of human rights. *The Annals of the American Academy of Political and Social Science*, 506(1):10–16, 1989.

[65] Lord Hoffmann. The universality of human rights. *Judicial Studies Board Annual Lecture*, 19(03), 2009.

[66] House of Representatives of the Philippines. Philippines house bill no.7396: Act establishing the artificial intelligence development authority, 2024. Proposes the creation of AIDA, including regulation of AI and complaint mechanisms for affected individuals.

[67] Geraint Howells and Stephen Weatherill. *Consumer protection law*. Routledge, 2017.

[68] Raphaële Xenidis Ivana Bartoletti. Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination, August 2023.

[69] Thorsten Jelinek, Wendell Wallach, and Danil Kerimi. Policy brief: the creation of a g20 coordinating committee for the governance of artificial intelligence. *AI and Ethics*, 1(2):141–150, 2021.

[70] Anna Jobin, Marcello Ienca, and Effy Vayena. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399, 2019.

[71] Isak Karabegović, Ermin Husak, Edina Karabegović, and Mehmed Mahmić. China is the leading country in the world in the implementation of robotic technology as the basic technology of industry 4.0. In *International Conference on Machine and Industrial Design in Mechanical Engineering*, pages 612–621. Springer, 2024.

[72] Kozo Kawai and Madoka Shimada. Merger control in japan: Select jurisdictional, procedural and substantive developments. In *Research Handbook on Global Merger Control*, pages 515–536. Edward Elgar Publishing, 2023.

[73] Andrew J Keith. Governance of artificial intelligence in southeast asia. *Global Policy*, 15(5):937–954, 2024.

[74] Christoph Kern, Ruben L. Bach, Hannah Mautner, and Frauke Kreuter. Fairness in algorithmic profiling: A german case study. *arXiv preprint arXiv:2108.04134*, 2021. Analysis of employment-risk profiling for job seekers in Germany, highlighting fairness disparities and the need for audit mechanisms.

[75] Els J Kindt. Privacy and data protection issues of biometric applications. *A Comparative Legal Analysis*, 12, 2013.

[76] Mercy King'ori. High court of kenya halts huduma namba rollout pending data protection compliance. *Future of Privacy Forum Blog*, 2022. Based on Republic v. Mucheru, Katiba Institute, 14 Oct 2021 ruling.

[77] Haksoo Ko, John Leitner, Eunsoo Kim, and Jonggu Jeong. Structure and enforcement of data privacy law in south korea. *International Data Privacy Law*, 7(2):100–114, 2017.

[78] Korea Communications Commission. 2024 work plan, 22 March 2024.

[79] Souichirou Kozuka. Japan's response to new technologies: Draft artificial intelligence research and development guidelines for international discussions. *Zeitschrift für Japanisches Recht*, 23(46):3–18, 2018.

[80] Isabel Kusche. Possible harms of artificial intelligence and the eu ai act: fundamental rights and risk. *Journal of Risk Research*, pages 1–14, 2024.

[81] Charles Labrecque. Asean issues guidelines for artificial intelligence, March 6th 2024.

[82] Filippo Lancieri. Narrowing data protection's enforcement gap. *Me. L. Rev.*, 74:15, 2022.

[83] Margaret Law. Reduce, reuse, recycle: Issues in the secondary use of research data. *IASSIST Quarterly*, 29(1):5–5, 2006.

[84] David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Michael Katell, and Morgan Briggs. Artificial intelligence, human rights, democracy, and the rule of law: a primer. *arXiv preprint arXiv:2104.04147*, 2021.

[85] Francesco Paolo Levantino and Federica Paolucci. Advancing the protection of fundamental rights through ai regulation: How the eu and the council of europe are shaping the future. *European Yearbook on Human Rights 2024*, 2024.

[86] Ross Levine. Law, endowments and property rights. *Journal of Economic Perspectives*, 19(3):61–88, 2005.

[87] Jing Li and Qinyuan Li. Data security and risk assessment in cloud computing. In *ITM Web of Conferences*, volume 17, page 03028. EDP Sciences, 2018.

[88] Zhi Li, Wenyi Zhang, Hengtian Zhang, Ran Gao, and Xingdong Fang. Global digital compact: A mechanism for the governance of online discriminatory and misleading content generation. *International Journal of Human–Computer Interaction*, 41(2):1381–1396, 2025.

[89] Fan Liang, Vishnupriya Das, Nadiya Kostyuk, and Muzammil M Hussain. Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet*, 10(4):415–453, 2018.

[90] Merlyna Lim. From activist media to algorithmic politics: The internet, social media, and civil society in southeast asia. In *Routledge Handbook of Civil and Uncivil Society in Southeast Asia*, pages 25–44. Routledge, 2023.

[91] Sun Sun Lim and Gerry Chng. Verifying ai: will singapore's experiment with ai governance set the benchmark? *Communication Research and Practice*, 10(3):297–306, 2024.

[92] Susie Lindsay and Thomas Nye. Human rights ai impact assessment backgrounder. *Law Commission of Ontario*, March 2025.

[93] Hu Moda Liu Xianglin. The second asian human rights forum reaches consensus: Technological development must have the protection of human rights as its fundamental guiding principle., 27 April 2025. Accessed October 10, 2025.

[94] Paul Luehr and Brandon Reilly. Data minimisation: A crucial pillar of cyber security. *Cyber Security: A Peer-Reviewed Journal*, 8(3):243–254, 2025.

[95] Gianclaudio Malgieri and Cristiana Santos. Assessing the (severity of) impacts on fundamental rights. *Computer Law & Security Review*, 56:106113, 2025.

[96] Monique Mann and Tobias Matzner. Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data & Society*, 6(2):2053951719895805, 2019.

[97] Ivan Manokha. Surveillance, panopticism, and self-discipline in the digital age. *Surveillance and Society*, 16(2), 2018.

[98] Alessandro Mantelero. Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer law & security review*, 32(2):238–255, 2016.

[99] Alessandro Mantelero. *Beyond data: Human rights, ethical and social impact assessment in AI*. Springer Nature, 2022.

[100] Jill Marshall. *Personal freedom through human rights law?: Autonomy, identity and integrity under the European convention on human rights*. Brill, 2009.

[101] Bertin Martens. *How DeepSeek has changed artificial intelligence and what it means for Europe*. Bruegel, 2025.

[102] Elizabeth Ann Masiello. *Privacy implications of biometric surveillance: The destruction of anonymity*. PhD thesis, Wellesley College., 2003.

[103] Lorna McGregor, Daragh Murray, and Vivian Ng. International human rights law as a framework for algorithmic accountability. *International & Comparative Law Quarterly*, 68(2):309–343, 2019.

[104] Media Diversity Institute. Confronting ai bias in southeast asia: Safeguarding democracy in the age of automation, 2024. Accessed: 2025-08-20.

[105] Susy Mendoza. Gdpr compliance-it takes a village. *Seattle UL Rev.*, 42:1155, 2018.

[106] Jacob Metcalf, Ranjit Singh, Emanuel Moss, Emnet Tafesse, and Elizabeth Anne Watkins. Taking algorithms to courts: A relational approach to algorithmic accountability. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 1450–1462, 2023.

[107] Sara Migliorini. China's interim measures on generative ai: Origin, content and significance. *Computer Law & Security Review*, 53:105985, 2024.

[108] Ministry of Science and Technology of the People's Republic of China. Opinions on strengthening the governance of science and technology ethics, 2022.

[109] Ministry of Science and Technology, Socialist Republic of Vietnam. Decision no. 1290/qd-bkhcn on principles for responsible ai development, 2024. Issued June 2024, establishing principles for human-centred and responsible AI systems.

[110] Fabio Morandín-Ahuerma. Ten unesco recommendations on the ethics of artificial intelligence. 2023.

[111] Xuelin Mu. Asian forum on human rights wraps up with chongqing consensus, April 2025. Accessed October 9, 2025.

[112] Vincent C Müller. Deep opacity undermines data protection and explainable artificial intelligence. *Overcoming opacity in machine learning*, 18, 2021.

[113] Prof Dr Eng Mutambara. Strategic framework for ai deployment. In *Deploying Artificial Intelligence to Achieve the UN Sustainable Development Goals: Enablers, Drivers and Strategic Framework*, pages 245–267. Springer, 2025.

[114] Eda Nano and Félix Tréguer. "'predictive' policing in france: Against opacity and discrimination, why a ban is needed", May 2025. Prepared as part of the Technopolice action-research series.

[115] National Assembly of the Republic of Korea. Artificial intelligence basic act and ai framework act, 2024. Enacted in December 2024, promulgated in January 2025, entering into force January 2026.

[116] Joelle Danielle Ngo Ndjama and Johan Van Der Westhuizen. Harnessing the power of responsible artificial intelligence for enhanced digital education leadership in higher education. In *Digital Leadership for Sustainable Higher Education*, pages 155–190. IGI Global Scientific Publishing, 2025.

[117] NITI Aayog, Government of India. Responsible ai for all: Principles for ethical and inclusive artificial intelligence, 2021. National framework emphasizing inclusivity, fairness, and non-discrimination in AI.

[118] Ana Brian Nougrères. Principles of transparency and explainability in the processing of personal data in artificial intelligence, 2023. UN Doc. A/78/310, August 16, 2023.

[119] OECD. Recommendation of the council on artificial intelligence, oecd/legal/0449. *OECD Legal Instruments*, 2025.

[120] Ministry of Foreign Affairs of the People's Republic of China. Ai capacity-building action plan for good and for all, September 2024. Accessed October 9, 2025.

[121] The Ministry of Industry, Information Technology of China, and sixteen other departments. Notice on issuing the implementation plan for the 'robotics plus' application initiative., 19 January 2023. Accessed October 10, 2025.

[122] Office of the High Commissioner for Human Rights. Access to remedy in cases of business-related human rights abuse: An interpretive guide, 18 October 2024.

[123] Office of the Privacy Commissioner of Canada. Privacy commissioners recommend halting clearview ai's facial recognition deployments in canada and deleting stored data, 2021. Prescribes cessation of collection and deletion of biometric data.

[124] Oh Byung-il. [commentary] regrettable passage of ai basic law in the national assembly that focuses on industry and ignores human rights, 27 December 2024.

[125] Kieron O'Hara, Nigel Shadbolt, and Wendy Hall. A pragmatic approach to the right to be forgotten. 2016.

[126] Anna Oosterlinck. Informal consultation with stakeholder. independent international scientific panel on artificial intelligence (ai) and global dialogue on a governance, 18 February 2025.

[127] Conor O'Mahony. There is no such thing as a right to dignity. *International Journal of Constitutional Law*, 10(2):551–574, 2012.

[128] Do Hyun Park, Eunjung Cho, and Yong Lim. A tough balancing act–the evolving ai governance in korea. *East Asian Science, Technology and Society: An International Journal*, 18(2):135–154, 2024.

[129] European Parliament and the Council. Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828, (artificial intelligence act), 2024.

[130] Parliament of Japan. Act on the promotion of research and development and the utilization of ai-related technologies, 4 June 2025.

[131] Rik Peeters and Arjan C Widlak. Administrative exclusion in the infrastructure-level bureaucracy: The case of the dutch daycare benefit scandal. *Public Administration Review*, 83(4):863–877, 2023.

[132] People Matters Global. Is algorithmic bias hurting southeast asia?, 2023. Accessed: 2025-08-20.

[133] Carolina Perlingieri et al. Responsabilità civile e robotica medica. 2018.

[134] Lourdes Peroni and Alexandra Timmer. Vulnerable groups: The promise of an emerging concept in european human rights convention law. *International journal of constitutional law*, 11(4):1056–1085, 2013.

[135] Yulu Pi and Maddie Proctor. Toward empowering ai governance with redress mechanisms. In *Cambridge Forum on AI: Law and Governance*, volume 1, page e24. Cambridge University Press, 2025.

[136] Bama Andika Putra. Governing ai in southeast asia: Asean's way forward. *Frontiers in Artificial Intelligence*, 7:1411838, 2024.

[137] K.s. Puttaswamy v. Union of India. Supreme Court of India, AIR 2017 SC 4161, recognizing privacy as a fundamental right.

[138] Usha Ramanathan. Exclusion by design: Aadhaar, biometric authentication and welfare exclusion in india. *Economic & Political Weekly*, 55(15):12–16, 2020.

[139] Filippo A Raso, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Levin Kim. Artificial intelligence & human rights: Opportunities & risks. *Berkman Klein Center Research Publication*, (2018-6), 2018.

[140] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R Varshney, Amit Dhurandhar, and Richard Tomsett. Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction*, 6(CSCW1):1–22, 2022.

[141] Protection Regulation. General data protection regulation. *Intouch*, 25:1–5, 2018.

[142] Huw Roberts, Josh Cowls, Jessica Morley, Mariarosaria Taddeo, Vincent Wang, and Luciano Floridi. The chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. In *Ethics, governance, and policies in artificial intelligence*, pages 47–79. Springer, 2021.

[143] Marc Rotenberg. Framework convention on artificial intelligence and human rights, democracy and the rule of law (council eur.). *International Legal Materials*, 64(3):859–902, 2025.

[144] Marc Rotenberg, Jeramie Scott, and Julia Horwitz. *Privacy in the modern age: The search for solutions*. New Press, The, 2015.

[145] Johnny Ryan. Global use of data and european responses. *BUSINESS AND POLICY CHALLENGES OF GLOBAL UNCERTAINTY: European Perspectives*, pages 299–310, 2025.

[146] Marietje Schaake. *The Tech Coup: How to Save Democracy from Silicon Valley*. Princeton University Press, Princeton, NJ, 2024.

[147] Bart Schermer. Risks of profiling and the limits of data protection law. In *Discrimination and privacy in the information society: Data mining and profiling in large databases*, pages 137–152. Springer, 2013.

[148] ASEAN Secretariat. Asean guide on ai governance and ethics, 2024.

[149] Cabinet Secretariat. Social principles of human centric ai (2019). 2019.

[150] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*)*, pages 59–68. ACM, 2019.

[151] Farida Shaheed. Report on the right to education and artificial intelligence, October 2024. UN Doc. A/79/520.

[152] Pawan Singh. Aadhaar and data privacy: biometric identification and anxieties of recognition in india. *Information, Communication & Society*, 24(7):978–993, 2021.

[153] Lu Siqi. Basis and path of corporate social responsibility governance ofgenerative artificial intelligence enterprises. *China Business and Market*, 37(12), 2023.

[154] Bernd Carsten Stahl, Laurence Brooks, Tally Hatzakis, Nicole Santiago, and David Wright. Exploring ethics and human rights in artificial intelligence–a delphi study. *Technological Forecasting and Social Change*, 191:122502, 2023.

[155] Elizabeth Stoycheff, Juan Liu, Kai Xu, and Kunto Wibowo. Privacy and the panopticon: Online mass surveillance's deterrence and chilling effects. *New media & society*, 21(3):602–619, 2019.

[156] European Data Protection Supervisor. Opinion 44/2023 on the proposal for artificial intelligence act in the light of legislative developments, 23 October 2023.

[157] Araz Taeihagh. Governance of artificial intelligence. *Policy and society*, 40(2):137–157, 2021.

[158] Margareth Theresia. Newly enacted law sets basis for nat'l development of ai. *Available at https://www. korea. net/NewsFocus/policies/view? articleId= 264071 (last accessed: 2025/01/23)*, 2024.

[159] Melodie Tieleman. Fairness in tension: A sociotechnical analysis of an algorithm used to grade students. *Cambridge Forum on AI: Law and Governance*, 1:e19, 2025. Examines fairness tensions in the Ofqual A-level grading algorithm.

[160] Yuichiro Tsuji. Gps investigations under constitution of japan–comparison with the us cases. *International and Comparative Law Review*, 18(1):179–197, 2018.

[161] UNESCO. Recommendation on the ethics of artificial intelligence, Adopted: 23 November 2021.

[162] United Nations General Assembly. Convention on the elimination of all forms of discrimination against women. GA Res 34/180, adopted 18 December 1979, entered into force 3 September 1981, 1979.

[163] United Nations Human Rights Council (Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance). Artificial intelligence and racial discrimination. Technical Report A/HRC/56/68, United Nations Human Rights Council, June 2024.

[164] Bart Van der Sloot. Legal fundamentalism: Is data protection really a fundamental right? In *Data protection and privacy:(in) visibilities and infrastructures*, pages 3–30. Springer, 2017.

[165] Dorine Eva Van Norren. The ethics of artificial intelligence, unesco and the african ubuntu perspective. *Journal of Information, Communication and Ethics in Society*, 21(1):112–128, 2023.

[166] Michael Veale and Frederik Zuiderveen Borgesius. Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4):97–112, 2021.

[167] Plixavra Vogiatzoglou. The ai act national security exception. *Verfassungsblog*, 2024.

[168] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International data privacy law*, 7(2):76–99, 2017.

[169] Jeremy Waldron. *Dignity, rank, and rights*. Oxford University Press, 2012.

[170] Lorna Woods. The uk's approach to regulation of digital platforms. In *Perspectives on Platform Regulation*, pages 329–350. Nomos Verlagsgesellschaft mbH & Co. KG, 2021.

[171] Fei Wu, Cewu Lu, Mingjie Zhu, Hao Chen, Jun Zhu, Kai Yu, Lei Li, Ming Li, Qianfeng Chen, Xi Li, et al. Towards a new generation of artificial intelligence in china. *Nature Machine Intelligence*, 2(6):312–316, 2020.

[172] Themis Dimitra Xanthopoulou, Nicole Tong, Rachele Carli, Maja Fjaestad, and Virginia Dignum. Dimensions of ai governance: A framework for evaluating global initiatives. In *to appear*. Springer, 2025.

[173] Karen Yeung. Algorithmic regulation: A critical interrogation. *Regulation & governance*, 12(4):505–523, 2018.

[174] Karen Yeung. Recommendation of the council on artificial intelligence (oecd). *International legal materials*, 59(1):27–34, 2020.

[175] Esmat Zaidan and Imad Antoine Ibrahim. Ai governance in a complex and rapidly changing regulatory landscape: A global perspective. *Humanities and Social Sciences Communications*, 11(1), 2024.

[176] Paolo Zatti. Note sulla semantica della dignità. *Maschere del diritto volti della vita*, pages 24–49, 2009.

[177] Frederik Zuiderveen Borgesius. Ai and discrimination: Prohibitions, challenges, and policy responses. *Computer Law & Security Review*, 41:105532, 2021.

[178] Valdemar Švábenský, Mélina Verger, Maria Mercedes T. Rodrigo, Clarence James G. Monterozo, Ryan S. Baker, Miguel Zenon Nicanor Lerias Saavedra, Sébastien Lallé, and Atsushi Shimada. Evaluating algorithmic bias in models for predicting academic performance of filipino students. 2024.

A Appendix: International Frameworks – Extended Texts and Details

A.1 United Nations Special Rapporteurs

- **Special Rapporteur on Human Rights while Countering Terrorism** Report A/HRC/54/21 (July 2023) by Fionnuala Ní Aoláin raised concerns on AI-enabled surveillance of journalists and activists, urging a moratorium until adequate safeguards for data protection and freedom of expression are in place [8].
- **Special Rapporteur on the Right to Privacy** Report A/78/310 (August 2023) by Ana Brian Nougrères: "Principles of Transparency and Explainability in the Processing of Personal Data in Artificial Intelligence" [105]. Report A/HRC/55/41 (2025) addressed neurodata and neurotechnologies, reiterating risks of opaque AI decision-making.
- **Special Rapporteur on the Right to Education** Report A/79/520 (October 2024) by Farida Shaheed on AI in education highlighted opportunities (inclusion, disability support) but also risks (educational disparities, alienation of teachers) [131].

A.2 ICESCR – Article 15(b)

Full text extract: "The States Parties to the present Covenant recognize the right of everyone to enjoy the benefits of scientific progress and its applications." (ICESCR, Art. 15(b)) [74].
Commentary: - Interpreted as requiring states to remove barriers to access technological advances, including AI. - Imposes obligations of transparency, documentation, oversight, and remedies in AI deployment [119].

A.3 OHCHR Reports and Guidance

- OHCHR reports on biometric surveillance, facial recognition, and predictive policing: risks of racial/social bias, opacity, and black-box effects [3,20].
- 2024 interpretative guidance on UNGPs, applying due diligence to AI lifecycle, stressing discrimination and data protection risks [107].
- Global Digital Compact (draft, 2025): - Proposes **Independent International Scientific Panel on AI**. - Global Dialogue on AI Governance [55,78].

A.4 OECD AI Principles – Full List

Five Principles (2019):

1. Inclusive growth, sustainable development, and well-being.
2. Human-centred values and fairness.
3. Transparency and explainability.
4. Robustness, security, and safety.
5. Accountability.

Policy Recommendations:

1. Invest in trustworthy AI R&D.
2. Foster an enabling AI ecosystem.

3. Ensure a sound policy environment.
4. Build human capacity and labour market transition measures.
5. Foster international co-operation.

2024 Updates: - Added environmental sustainability. - Strengthened accountability to include bias, labour rights, intellectual property. - Transparency reframed as contestability of algorithmic decisions.

A.5 G20 AI Guidelines – Text Extract

Adopted at the Osaka Summit (2019): - Fairness, transparency, accountability, privacy, and rule of law. - High-level political declaration, without monitoring/enforcement [62,124]. Unlike OECD principles, lacks operational detail; functions as a diplomatic alignment tool.

A.6 Global Partnership on AI – Founding Details

Launch: June 2020 (proposed at 2018 G7, hosted by OECD). **Membership:** 20+ states including EU members, Canada, Japan, Korea.
Focus areas: - Responsible AI, Data Governance, Future of Work, Innovation and Commercialization. - Multi-stakeholder structure (states, civil society, academia, industry).
Core normative base: OECD AI Principles, UNGPs.

A.7 G7 Hiroshima Process – Full List of Principles

1. Risk management across lifecycle.
2. Incident response mechanisms.
3. Transparency: public reporting of capabilities and limitations.
4. Information-sharing on incidents.
5. Risk-based governance.
6. Strengthened physical, cyber, and insider security.
7. Content authentication (traceability, provenance).
8. Research prioritisation on risk mitigation.
9. Address major challenges (climate, education, health).
10. Support international technical standards.
11. Strengthen personal data and IP protection.

These are voluntary but influential in shaping national codes of conduct.

A.8 UNESCO Recommendation on the Ethics of AI – Extracts

Adopted: November 2021, 194 Member States [139].
Key Principles:

- Human dignity at the centre [98].
- Inclusivity, gender equality, and diversity [103].
- Environmental sustainability [143].
- Education and training for responsible AI use.

Limitations: - Voluntary. - Preventive approach, limited to pre-deployment phases. - Challenges for adaptive/self-learning AI systems.

A.9 IEEE Ethically Aligned Design – Extended Details

The IEEE's *Ethically Aligned Design* (EAD) principles, developed through the Global Initiative on Ethics of Autonomous and Intelligent Systems, provide one of the most comprehensive voluntary frameworks for embedding ethics in AI systems.

Key principles:

- Promote and protect rights to life, safety, privacy, equality, and freedom of expression.
- Prevent discrimination based on race, gender, religion, disability, sexual orientation, or other characteristics.
- Ensure human oversight and agency, minimising risks of manipulation and coercion.
- Strengthen accountability by enabling outcomes to be traced back to responsible actors.
- Build public trust through transparency and verifiable accountability mechanisms.

Implementation challenges:

- Embedding ethical safeguards requires advanced tools such as algorithmic audits, adversarial testing, and ethical risk modelling, which remain difficult to operationalise at scale.
- Establishing diverse, interdisciplinary ethics review boards is resource-intensive and inconsistently adopted across organisations.
- Voluntary status means no enforcement; market and efficiency pressures often outweigh adoption.

The EAD initiative has nonetheless been influential in both industry and academic settings, serving as a reference point for operationalising human rights in technical design processes. It complements normative frameworks such as the OECD AI Principles and UNESCO Recommendation, by targeting developers and engineers as key implementers of ethical AI.

A.10 Raoul Wallenberg Institute – Extended Details

The Raoul Wallenberg Institute of Human Rights and Humanitarian Law (RWI) is a multidisciplinary research and policy institute dedicated to advancing human rights. It conducts applied research, provides education, and engages in policy dialogue, including on the human rights impacts of emerging technologies.

Key focus areas in relation to AI:

- Investigating algorithmic bias and its consequences for equality and non-discrimination.
- Addressing accountability gaps in AI governance and promoting transparent design.
- Exploring AI's role in healthcare, justice, public safety, and social welfare, with a human dignity–centred approach.
- Promoting inclusive, multi-stakeholder participation in AI governance.

Contribution: RWI's work is advisory and educational rather than regulatory. It strengthens the conceptual foundations of rights-based AI governance and supports policymakers, civil society, and academia in understanding both risks and opportunities.

Although not binding, its influence lies in building human rights capacity and shaping debates around ethical and inclusive AI development.

B Appendix: Human Rights and AI at the Regional Level: Asia – Extended Texts and Details

B.1 Approach to AI regulation in South Korea – Extended Details

South Korea is increasingly adopting AI across sectors like healthcare, education, and public administration. AI is used in predictive healthcare models, personalized learning, and smart city initiatives, aiming to improve services and quality of life. However, rapid AI development raises significant societal concerns. Key issues include algorithmic bias, especially in areas like recruitment and criminal justice, and the potential misuse of personal data in AI systems, raising privacy and transparency concerns. Additionally, there are fears of job displacement due to automation, leading to social instability. These concerns have sparked debates on the need for strong ethical guidelines to ensure AI respects human rights. South Korea's government and civil society are calling for greater transparency, accountability, and human rights impact assessments (HRIAs) to address the risks and build public trust in AI technologies.

Overall, South Korea's AI regulatory framework does not impose overly burdensome requirements on the industry, offering companies considerable flexibility in the development and deployment of AI technologies. The hard regulatory measures include: (a) existing legislative frameworks, such as the Framework Act on Informatisation and the Personal Information Protection Act ("PIPA"); (b) new legislation, notably the Act on the Development of Artificial Intelligence and Establishment of Trust (the "AI Basic Act"); (c) and the proposed Act on the Protection of Artificial Intelligence Service User [78].

B.2 Approach to AI regulation in Japan – Extended Details

Japan, as a nation at the forefront of addressing mature society challenges — including declining birthrate, ageing population, labour shortage, and rising fiscal spending — views AI as a core technology to tackle these issues, advance the UN's Sustainable Development Goals (SDGs), and underpin its "Society 5.0" initiative. While AI offers significant societal benefits, its profound impact requires prudent development; thus, Japan seeks to shift to an "AI-Ready Society" via comprehensive reforms to social systems, industrial structures, and governance. Though AI lacks a clear universal definition, there is broad consensus on identifying core AI technologies—often integrated into complex information systems— and principles for AI are tailored to such systems. Ultimately, successful AI governance and realization of Society 5.0 depend on close collaboration among all stakeholders [149].

B.3 Approach to AI regulation in China – Extended Details

At present, China is undergoing a crucial phase of digital transformation. The new wave of technological revolution and industrial transformation, represented by information technologies such as artificial intelligence, block-chain, and big data, has become a significant driving force for China's economic and social development. Uniform legislation on artificial intelligence often requires a long period of time. In the current stage, China has adopted a decentralized legislative approach for different scenarios to meet the needs of rapid development of artificial intelligence. In the realm of artificial intelligence governance, the Ethical Guidelines for the New Generation of Artificial Intelligence (2021) [171] and the Opinions on Strengthening the Governance of Science and Technology Ethics (2022) [108] pioneered the establishment of a soft law framework characterized by an 'ethics-first' approach. Building upon this foundation, a series of departmental regulations — including the Provisions on the Administration of Algorithmic Rec-

ommendation in Internet Information Services (2021) [23], the Provisions on the Administration of Deep Synthesis in Internet Information Services (2022) [58], and the Interim Measures for the Administration of Generative Artificial Intelligence Services (2023) [57]. On May 29, 2025, the Artificial Intelligence Subcommittee of the National Committee on Science and Technology Ethics formally issued the Ethical Review Guidelines for Generative AI Algorithms. For the first time, this regulatory document requires enterprises to submit ethical impact assessment reports during the algorithm training phase. It specifically stipulates requirements regarding the legitimacy of data sources, mechanisms for bias mitigation, and the boundaries for applying deepfake technology. These departmental regulations translate ethical imperatives into binding compliance obligations, explicitly requiring enterprises to establish mechanisms for content generation labelling, lawful review of training data, algorithmic fairness assessment, and false information prevention. They are underpinned by a comprehensive liability framework spanning administrative penalties to criminal prosecution, systematically reinforcing enterprises' responsibility to respect human rights throughout technological development and operational processes. In specific risk domains, the Provisions on the Administration of Security Vulnerabilities in Network Products [56] in Network Products establish a dual-constraint mechanism comprising mandatory vulnerability reporting coupled with joint disciplinary measures for dishonest conduct, strictly prohibiting enterprises from exploiting security vulnerabilities for profit. With the acceleration of the algorithm era, the challenges of defining and assigning algorithmic responsibility have become increasingly prominent, raising public concerns about the implications of widespread algorithm use.

Furthermore, the Measures for the Review of Science and Technology Ethics (for Trial Implementation) (2023 Ethical Review Measures) issued on 7 September 2023 further clarified the ethical review mechanism for science and technology, encompassing risk assessment, prevention and control, follow-up monitoring, and remedial measures. In terms of content review, it focused on human rights risks in areas such as personal information protection, the right to know, and special safeguards for vulnerable groups. Recently, the Measures for the Management and Service of Artificial Intelligence Ethics (for Trial Implementation) (Public Consultation Draft) was released for public consultation from 22 August to 22 September. The draft inherits the four review procedures of "general, simplified, expert review, and emergency" in 2023 Ethical Review Measures, ensuring the unity and connection with the national scientific and technological ethics governance system. Meanwhile, based on the uniqueness of artificial intelligence technology, AI Ethics Measures identifies specific ethical principles and compliance risks in the field of AI, specifically regulates exclusive ethical issues such as data, algorithms, and automated decision-making in the AI domain, and proposes the establishment of a service centre to provide supporting guidance and supervision for AI ethics compliance.

In summery, China's approach to AI governance prioritizes both technological development and risk control. By adopting a strategy of "decentralised legislation and scenario-specific regulation", China has established a multi-layered governance framework. This framework is anchored by foundational laws and detailed through specialized regulations. However, the effectiveness of this governance faces challenges, including potential regulatory overlaps or gaps due to multi-agency oversight, difficulties in balancing transparency requirements in practice, and a still-underdeveloped attribution and remedy mechanism tailored to the unique characteristics of algorithmic technology. Collectively, these features and challenges shape China's distinctive model, which seeks to balance social stability, economic development, and human rights protection amidst its rapid advancement and application of AI technologies.

B.4 Developments in AI Governance Legislation in Other Asian Countries – Details

Table 5: National Level Progress on AI Governance and Human Rights in Asia

Category	Countries	Law/Policy	Key features
Human Rights Explicitly Referenced	Thailand	Digital Thailand — AI Ethics Guideline	Emphasizes alignment with laws, ethics, and human rights. Highlights six core principles, including transparency, accountability, and fairness.
	Vietnam	Decision No. 1290/QD-BKHCN	Prioritizes human-centered AI development, stressing respect for human rights and dignity. Focuses on preventing bias and unfairness in AI systems.
	Indonesia	Circular Number 9 of 2023	Stresses ethical values in AI utilization, including humanity, inclusivity, security, transparency, and accountability. The National Artificial Intelligence Strategy identifies priority sectors for AI application.
	Philippines	Joint Memorandum Circular of 18 April 2024, Bill No. 7396	Conforms to global AI standards ensuring alignment with human rights, well - being, and sustainable development
	Malaysia	National Guidelines on AI Governance & Ethics	Encourages voluntary adoption of seven AI principles alongside existing laws, focusing on human benefit and happiness
	India	Responsible AI for All	Integrates fundamental constitutional rights into AI governance, enabling fundamental rights in AI system design
	Australia	AI Ethics Principles	Voluntary principles guiding ethical AI design, deployment, and operation with five ”cornerstones” for AI assurance in government
	New Zealand	Public Service AI Framework	Envisions responsible AI adoption to modernise services and enhance citizen outcomes, based on OECD AI principles
No Comprehensive Framework or Reference to Human Rights	Kazakhstan	Draft law on artificial intelligence	Built on fairness, legality, accountability, and human well — being principles, prohibiting unauthorised data collection
	Pakistan	Currently no specific law	Draft encourages AI adoption, addresses specific application risks, and urges digital technology evaluation and civil law framework drafting

Table 5: National Level Progress on AI Governance and Human Rights in Asia

Category	Countries	Law/Policy	Key features
No Comprehensive Framework or Reference to Human Rights	Brunei	Guide on Artificial Intelligence Governance and Ethics	Promotes ethical and responsible AI development and adoption with seven principles including transparency, security, and fairness
	Cambodia	Ethics of AI Readiness Assessment	Assesses Cambodia's ethical AI development and use preparedness to inform national AI policy formulation
	Bangladesh	National Artificial Intelligence Policy 2024	Addresses social, legal, and ethical issues related to AI implementation across sectors, though lacks comprehensive guidelines in some key areas
	Singapore	AI Verify	AI governance testing framework and software toolkit validating AI performance across multiple principles like transparency and safety
	Mongolia	Digital Nation programme	Signifies growing AI governance ambition via strategic initiatives with UNDP — supported AI readiness assessments
	Laos	Currently no specific law	Actively explores AI integration into governance, industry, and ethics sectors
	Myanmar	Currently no specific law	Focuses on broad technology and communication aspects, not specifically addressing AI issues

B.5 Developments in AI Governance Legislation at a Regional Level – Details

Table 6: Regional Level Progress on AI Governance and Human Rights

Initiative/Treaty	Key Features
ASEAN Digital Masterplan 2025 (ADM 2025)	Aims to drive digital transformation in the ASEAN region and has taken steps to integrate AI governance.
ASEAN Guide on AI Governance and Ethics	Non-binding guide emphasizing core principles such as transparency, fairness, security, reliability, privacy, accountability, and human centricity. It stresses the need for countries to retain agency over AI-driven outcomes and to design and use AI systems to promote human well-being and protect individuals from harm.
ASEAN Responsible AI Roadmap (2025–2030)	Introduces a Readiness Assessment Framework to categorize countries based on their stage of AI governance development, allowing for differentiated support and benchmarking. It encourages governments to tailor strategies according to their institutional maturity and includes a direct question about national human rights protections in AI policies.
Chongqing Consensus	Adopted at the 2025 Asian Forum on Human Rights, it provides a regional perspective on aligning technological progress with human rights protection. It calls for regional leadership in shaping ethical responses to generative AI and other emerging technologies and encourages cooperation on equitable governance structures and cross-border capacity-building.

C Appendix: Full Illustrative Cases by Thematic Focus

C.1 Privacy and Data Protection — Expanded Cases

- **DeepSeek Case:** In Germany, the data protection commissioner ordered the removal of the Chinese chatbot DeepSeek from Google and Apple's app stores, due to what was considered an unlawful transfers of users' data in China [101]. Italy blocked the access to this service, due to the lack in China of GDPR equivalents safeguards for individuals' personal data and the Netherlands restricted the general use of the chatbot, specifically prohibiting its government use. This app has also been banned in Korea by the Personal Information Protection Commission, with its unblocking subject to the provider's ability to guarantee privacy compliance.
- **Predictive Travel Surveillance Case:** The EU Court of Justice expressed its negative opinion on the use of fully automated traveller risk assessment systems. The decision was made based on the case of a passenger who was blocked at Amsterdam airport by a predictive AI system based on PNR data. The Court emphasises the high risk to privacy and data protection posed by such profiling systems, which were used without any form of human oversight [48].
- **Generative AI Models Case:** The EU Data Protection Authorities have opened investigations into Open AI's generative models, raising doubts about their compliance with the GDPR. In particular, a negative opinion has been expressed towards providers and businesses that use these models, due to the inadequacy of the profiling system and the scraping of public data.
- **Meta Cases:** The Irish Data Protection Commission fined Meta €1.2 billion in 2023 for transferring user data from Europe to the United States, without ensuring the necessary safeguards for such cases, thereby violating the GDPR [44]. The case caused such a stir that it has set a precedent which could be used in the future against other big tech companies operating globally.
- **Amazon Case:** The National Commission for Data Protection fined Amazon a total of €746 million for tracking its users without their informed consent and without complying with the transparency requirements of the GDPR [29].
- **LinkedIn Case:** In 2024, Ireland issued a €310 million fine against LinkedIn for breaching its obligations of transparency and lawful justification in processing its users' data for behavioural advertising purposes [145].
- **Alibaba Cloud Case:** Alibaba Cloud was fined in China for failing to disclose a concerning cybersecurity vulnerability that revealed a related infringement of privacy and security obligations [87].
- **LINE Corporation Case:** The LINE messaging app was subjected to a thorough investigation after the Japanese competent authorities noticed a data leak to Chinese engineers, in total breach of the Japanese Act on the Protection of Personal Information [72].
- **Kominfo:** In 2023, the Indonesia Ministry of Communication and Information issued ethical guidelines requiring that AI use respect human values and include safeguards for personal data protection. However, as a non-binding instrument, the circular lacks enforcement mechanisms, raising concerns about its effectiveness in mitigating privacy risks from AI deployment [9].

C.2 Equality and Non-Discrimination — Expanded Cases

- **The Netherlands**: The child benefits scandal has become a landmark case of algorithmic governance risks, showing how opaque profiling in public administration can lead to systemic discrimination, lack of accountability, and severe social harm [131]. Similarly, the SyRI system for fraud detection was ruled unlawful by a Dutch court in 2020 for violating the right to privacy and the principle of non-discrimination, highlighting how algorithmic risk scoring without transparency and safeguards undermines human rights, raising questions of legality and proportionality under the European Convention on Human Rights (ECHR) [14].
- **France**: Predictive policing initiatives such as PAVED (used by the Gendarmerie Nationale) have been criticised for opacity and lack of transparency, with concerns that data-driven feedback loops may reinforce discriminatory policing, prompting calls for a complete ban on human rights grounds. [114]. In the absence of proper safeguards and oversight, these algorithms may entrench rather than reduce bias (Maviş 2023).
- **Germany**: The Federal Employment Agency's "Arbeitsmarktchancen-Index" profiles job seekers into categories of high, medium, or low employment prospects, determining the support they receive. The system has been criticised for opacity and reliance on sensitive variables such as age, health, and migration background, raising concerns of indirect discrimination and unequal access to social rights [74].
- **United Kingdom**: The 2020 *Ofqual* grading algorithm, used to assign A-level results during the COVID-19 pandemic, disproportionately downgraded students from disadvantaged schools, prompting public outcry and highlighting risks of systemic discrimination, lack of transparency, and denial of equal access to education [159].
- **China:** Pilot projects for social credit and credit-scoring systems have raised concerns of indirect discrimination, as proxies such as geographic location, social networks, or socio-economic markers can systematically disadvantage certain groups [89].
- **India:** The Aadhaar biometric identification system, when used for welfare authentication, has resulted in the exclusion of vulnerable groups such as the rural poor, women, and the elderly from essential entitlements, highlighting how large-scale digital infrastructures can reinforce structural inequalities [138].
- **South Korea:** AI-based hiring tools have been criticised for opacity and discriminatory outcomes, while the AI chatbot *Lee Luda* drew attention for biased and offensive speech. Ride-hailing platforms have been investigated for algorithmic manipulation disadvantaging certain drivers, and immigration systems faced scrutiny for unauthorised sharing of facial data for AI training [47].
- **Indonesia:** Job-matching platforms using AI have been found to disadvantage female applicants due to historical data bias, reflecting systemic occupational segregation in training datasets [132].
- **Philippines (Credit Scoring):** AI-based financial services for credit evaluation risk excluding vulnerable groups, as models trained on Western-centric datasets fail to capture local demographic and linguistic realities [104].
- **Philippines (Education):** Learning analytics systems deployed in higher education (e.g. Canvas LMS) have been studied for bias in predicting student performance. While one major audit found no group bias, the case illustrates the need for fairness audits in educational AI [178].

C.3 Remedies and Access to Justice — Expanded Cases

- **Netherlands – SyRI case (2020):** The District Court of The Hague annuled the System Risk Indication welfare fraud detection tool (SyRI), not only due to its discriminatory effects but because individuals had no way to challenge opaque algorithmic risk scores. The judgment foregrounded contestability and transparency as procedural remedies [14, 8].
- **France – Conseil d'État (2018):** France's highest administrative court ruled that individuals affected by algorithmic decision-making in public administration have a right to access the "rules and characteristics" of the algorithm. This decision established an important precedent in judicially enforced procedural remedy by ensuring transparency and contestability [33].
- **United Kingdom – ICO enforcement actions (2019–2021):** The UK Information Commissioner's Office has used its regulatory powers to impose corrective measures on companies deploying opaque or unfair AI-driven credit scoring systems. This demonstrates how administrative enforcement can function as a structural remedy to protect rights [170].
- **India – Aadhaar litigation (2018, *Puttaswamy v. Union of India*):** The Supreme Court curtailed the mandatory use of Aadhaar biometric authentication in certain welfare contexts, citing the need to safeguard due process and ensure that individuals have access to redress when excluded from essential services [152].
- **South Korea – Constitutional Court on communications surveillance (2018):** The Court restricted government bulk metadata collection, finding that the absence of transparency and avenues for individual redress violated constitutional rights. This decision reinforced judicial remedies against opaque state surveillance practices [77].
- **Japan – Supreme Court on GPS surveillance (2017):** The Court ruled that warrantless GPS tracking by police infringed the right to privacy. The judgment underscored the need for clear procedural safeguards and judicial oversight as remedies in cases involving new technologies [160].
- **China – Personal Information Protection Law enforcement (2021) :** Chinese courts have begun hearing civil claims against companies for unlawful AI-driven data processing, such as facial recognition cases against shopping malls. These early rulings illustrate how statutory remedies and private enforcement mechanisms are emerging in the AI context [45].
- **Indonesia – Job-matching platforms (2023):** Following complaints of gender bias in AI-driven recruitment, regulators required corrective audits and transparency reporting. This shows how administrative oversight can provide sector-specific remedies in employment contexts [132]. Civil society and media advocacy played an important role in pressuring for these interventions, reflecting the broader role of algorithmic politics in Southeast Asia [90].
- **Philippines – Credit scoring oversight (2024):** The National Privacy Commission intervened against AI-based financial scoring models that excluded low-income groups, mandating fairness audits and redress procedures. This illustrates the role of regulators in creating collective remedies in financial services [104].
- **Xiamen Court Ruling (2025):** The Xiamen Maritime Court in China issued a precedent-setting judgment requiring litigation agents to disclose comprehensively any use of AI tools in judicial proceedings, including their scope, purpose, and data sources. This measure aims to safeguard transparency, accountability, and procedural fairness in trials involving automated systems. NEED REFERENCE FOR THIS CASE!