Policy Network on Artificial Intelligence (PNAI)
Discussion Paper
DRAFT VERSION, 1.10.2024

Liability for Harms Across the AI Lifecycle: Incentivizing Accountability

PNAI Sub-group on Liability as a mechanism for supporting AI accountability

1. Introduction

Policymakers and stakeholders focused on AI governance consistently highlight the need to hold AI developers and deployers accountable for harms caused across the AI lifecycle.[1] Yet, to date, mechanisms to ensure accountability remain poorly defined.[2] Many of the accountability frameworks under consideration in various jurisdictions and through intergovernmental agencies – for example, algorithmic impact assessments and audit-based monitoring – will have difficulty keeping pace with the quickly-morphing risks that will inevitably accompany AI's evolution.[3] Nevertheless, policymakers have an opportunity to stake a resilient approach to AI accountability, and to incentivize responsible AI development processes and outcomes, by establishing clear guidelines regarding **legal liability for harms**.[4] Liability can be a critical lever in mitigating AI-related risks ranging from algorithmic bias leading to discriminatory outcomes in hiring or lending, to AI-driven misinformation campaigns that can destabilize democracies, to malfunctions in AI-controlled critical infrastructure that can jeopardize public safety.[5]

With this discussion paper, we hope to progress the conversation about AI liability within the **global AI governance community**. While attention to AI liability has been prominent in the European Union for several years,[6] a globally coordinated approach to AI liability

[1] OECD, *Advancing accountability in AI*, Feb. 2023
[2] G. Noto La Diega & L.C.T. Bezerra, *Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive*, Sept. 2024
[3] Ibid.
[4] H. Zech, *Liability for AI: Public Policy Considerations*, Jan. 2021
[5] C. Wendehorst, *Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks*, 2022
[6] C. Novelli, F. Casolari., P. Hacker, G. Spedicato & L. Floridi, *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity*, March 2024

principles will incentivize safeguards across the digital divide, protecting individuals and communities worldwide from potential AI-related harms.[7] Liability frameworks fill critical gaps in AI governance as they keep pace with evolving risks, align incentives with responsible development and deployment, providing a path to accountability and recourse for damages.[8] Rather than advocating for a specific legislative model, we highlight scholarship and existing policy work – and advocate for a more prominent discussion of **principles** which can guide jurisdictions in responding to the complexities that accompany AI liability determinations.

2. Liability for AI systems - unique and urgent challenges

The universe of potential AI-related harms from facial recognition systems leading to wrongful arrests, to financial panic caused by faulty AI-driven market analyses, to discriminatory hiring based on AI-enabled human resource applications, to hazards caused by AI-led oversight of physical infrastructure such as water supply systems.[9] The **urgent need to incentivize AI developers and deployers** to proactively safeguard against such potential harms is both clear and immediate. Liability frameworks have helped create powerful economic incentives for innovative safeguards and risk mitigation strategies across industries as varied as food and beverage (e.g. improved food traceability and labeling for allergens), automotive (seatbelts, airbags, advanced driver assistance systems), and pharmaceutical industries (improved drug labeling and warning systems).[10] As is the case for these industries, liability frameworks pertaining to AI-related harm can be a critical component of a broader regulatory framework, and a force for safety innovations.

Liability frameworks offer **unique advantages** in AI governance.[11] As new risks emerge with advancing AI technologies, liability frameworks can naturally adapt to address these harms without requiring constant regulatory updates. This inherent scalability will make liability an indispensable tool in mitigating risks over time.[12]

However, legal experts argue that traditional liability frameworks are not fit-for-purpose for AI systems.[13] As AI systems' capacities progress autonomously, models produce outcomes that are not fully predictable – even to their creators. This autonomy blurs

[7] UN AI Advisory Body, *Governing AI for Humanity*, Sept. 2024
[8] G. Weil, *Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence*, June, 2024
[9] UN AI Advisory Body, *Governing AI for Humanity*, Sept. 2024
[10] C.M. Sharkey, *The Irresistible Simplicity of Preventing Harm*, July, 2023
[11] G. Weil, *Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence*, June, 2024
[12] M.H. Pfeiffer, *First Do No Harm: Algorithms, AI, and Digital Product Liability,* Sept. 2023
[13] H. Zech, *Liability for AI: Public Policy Considerations*, Jan. 2021

conventional lines of responsibility, complicating efforts to assign liability. The complexity will be further compounded by the web of contractual obligations and varied risk management approaches within the AI ecosystem. Amid these built-in complexities, ***policymakers and stakeholders have yet to come to a consensus regarding key questions on assigning liability within the AI lifecycle***.[14] For example, if an AI-powered medical diagnostic system misses a critical, treatable condition due to underlying biases in its training data, should liability be assigned to the healthcare provider, the medical application developer, to the underlying foundational system, or apportioned to some degree across these players?[15]

Given the countless, thorny, similar questions which will arise, it will be vitally constructive for global AI governance stakeholders to coordinate regarding **AI liability principles and standards**.[16] Neither the opacity, nor the autonomy of AI systems ought to exempt developers and deployers from accountability. In fact, the lack of transparency pertaining to these systems elevates the need to incentivize rigorous safeguarding, in part through ensuring companies will be held responsible for harms via clear liability frameworks. Without such mechanisms, damages caused by AI systems will be borne by faultless individuals, communities, and the public-at-large.[17]

2.1 Defining liability – and types of liability

Liability refers to the legal and financial responsibility for harm or damage caused by AI systems, encompassing obligations from developers and deployers to compensate affected parties for losses resulting across the AI lifecycle. Such harms can be incurred within the AI-training phase (e.g. web-scraping to train AI systems in a manner that sweeps up personal/private data or intellectual property) or pertaining to harmful AI outputs. Affected parties might include individuals, private organizations, or public entities.[18]

Our sub-group's work and this discussion paper focus on AI and ***product liability*** and ***civil liability***, leaving ***criminal liability*** out of scope.[19] Product liability is a legal concept that would hold developers and deployers responsible for harms caused by defects in the AI

[14] Center for Humane Technology, *A Framework for Incentivizing Responsible Artificial Intelligence Development and Use*, Sept. 2024
[15] W.N. Price II, S. Gerke, G. Cohen, *Potential Liability for Physicians Using Artificial Intelligence*, 2019
[16] C. Frattone, *Reasonable AI and Other Creatures. What Role for AI Standards in Liability Litigation?*, 2022
[17] UN AI Advisory Body, *Governing AI for Humanity*, Sept. 2024
[18] C. Wendehorst, *Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks*, 2022
[19] As a multistakeholder group representing diverse nations with vastly different criminal justice systems, the focus of our shared discussion has been on financial penalties versus *criminal liability*. While out of scope for this paper, criminal liability represents another potential avenue for addressing flagrant misconduct.

products or services they have made available to the public. Civil liability represents a broader legal category allowing not only individuals and organizations but states and Governments to seek remuneration for harms to protect public interests or recover damages on behalf of their citizens.[20] While product liability generally operates on a *strict liability* standard – requiring only proof of defect and resulting harm without a need to prove negligence – other civil liability mechanisms generally require proof of negligence or breach of duty.[21]

Administrative liability refers to penalties imposed by regulatory bodies or Government agencies for non-compliance with AI regulations.[22] For example, Article 99 of the EU AI Act[23] establishes administrative fines up to 35 million euros or 7% of global annual turnover for violation of prohibited practices including for example deploying subliminal techniques to exploit behavior; exploiting vulnerabilities of specific groups; certain kinds of AI-enabled biometric identification systems to monitor public; failure to engage comprehensive risk management systems; failure to use high quality, validated training data which have been thoroughly examined for biases; failure to maintain sufficient transparency that will allow proper evaluation of high-risk systems; failure to ensure ongoing human oversight of high-risk systems.[24]

We hope to encourage global policymakers to incentivize safeguarding and support accountability for AI developers and deployers by clarifying legal responsibilities and financial risks. Complex questions that will need to be navigated in developing clarity about liability for AI include: Who is financially responsible when an autonomous vehicle causes an accident? How should damages be apportioned if an AI trading algorithm causes panic in financial markets? Who bears financial responsibility if a medical diagnostic system produces skewed outcomes which harm patient health?[25]

2.2 The rationale for harmonizing frameworks

Coordinating AI liability principles on a global scale will incentivize developers and deployers to adhere to consistent standards, preventing "regulatory arbitrage" and "forum shopping" by companies seeking lenient jurisdictions, and leveling the playing field for

[20] European Parliamentary Research Service, *Proposal for directive on adapting non-contractual civil liability rules to artificial intelligence*, Sept. 2024
[21] H. Zech, *Liability for AI: Public Policy Considerations*, Jan. 2021
[22] A. Bertolini, *Artificial Intelligence and civil liability*, Jan. 2020
[23] European Union, *EU Artificial Intelligence Act, Article 9: Risk Management System*, 2024
[24] European Union, *EU Artificial Intelligence Act, Article 16: Obligations of Providers of High-Risk Systems*, 2024
[25] H. Zech, *Liability for AI: Public Policy Considerations*, Jan. 2021

deployment worldwide.[26] In order to make good on Global Digital Compact commitments to closing the digital divide, Global Majority countries which lack resources for comprehensive AI governance must benefit from the collective expertise and enforcement capabilities of the international AI governance community.[27] Harmonizing AI liability will help protect vulnerable individuals, communities and Global Majority countries from bearing the brunt of AI-related harms. Additionally, the questions surrounding AI liability across the value chain are intricate, challenging to navigate, and are likely to **cross national boundaries**. Harmonized liability principles can drive uniform requirements for AI/algorithmic transparency, making it easier for vulnerable individuals and communities to seek redress and recourse for harms, reducing the likelihood that those most at risk will be exploited or harmed, and increasing the likelihood the benefits AI offers are shared within a safer, more inclusive market ecosystem.[28]

2.3 AI-Specific Complexities for Liability Frameworks

The role of transparency and explainability. The opacity of AI systems, particularly large language models, pose significant challenges for any liability regime. To effectively assess potential *bias-related harms*, for example, it will be necessary to access valid indicators on the factors that contributed to decisions.[29] For liability frameworks to be meaningful and enforceable, jurisdictions will need access to information about how systems factored inputs and characteristics into their decisions – *algorithmic transparency* – a standard that has been thus far promised far more generously than it has been provided. Ideally, clearer and stronger liability frameworks will finally incentivize compliance with algorithmic transparency commitments, enabling regulators and adjudicators to determine whether AI systems are functioning in an unbiased manner.[30]

Liability Across the AI Lifecycle and Supply Chain. Harms can occur at various stages across the AI lifecycle, from development to deployment and ongoing use: during data collection (e.g. from the improper use of personal data and/or intellectual property), in the deployment phase, or as a result of the AI system's ongoing learning and adaptation.[31] These lifecycle complexities are compounded by interwoven activities within an AI supply

[26] G. Noto La Diega & L.C.T. Bezerra, *Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive*, Sept. 2024
[27] UN Office of the Secretary-General's Envoy on Technology, *Global Digital Compact*, Sept. 2024
[28] C. Wendehorst, *Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks*, 2022
[29] H. Zech, *Liability for AI: Public Policy Considerations*, Jan. 2021
[30] G. Noto La Diega & L.C.T. Bezerra, *Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive*, Sept. 2024
[31] R. Ashmore, R. Calinescu & C. Paterson, *Assuring the machine learning lifecycle: Desiderata, methods, and challenges*, 2021

chain involving data providers, model developers, the software companies that incorporate these AI models, practitioners, and end-users – all of which can make identifying the defects which cause harm unusually difficult.[32] Underlying biases in training data can interact with flaws in the life cycle (e.g. inadequate model training) or the supply chain to produce discriminatory outcomes.[33] To prevent downstream harms, liability frameworks might implement a greater emphasis at the source – a "chain of responsibility" approach – such that a greater onus is placed on foundational players to implement robust safeguards and quality controls.[34] An emphasis on the responsibilities of foundational companies will help incentivize more careful risk assessment of partners and providers downstream.

Adjudicating AI Liability. The "black box" nature of AI systems presents significant challenges in adjudicating liability cases, and the opacity of AI decision-making processes make it difficult for traditional courts to properly assess fault and causation. Some legal experts have argued that specialized courts or tribunals might be necessary, equipped with the technical expertise necessary to make informed decisions on liability.[35] More immediately, members of any judicial system adjudicating AI harms will need to be adequately educated about this technology's unique complexities.[36]

Indemnity, Contractual Liability and AI. In many industries involving significant risks, businesses use contracts to allocate responsibilities and liabilities – which might include *indemnification clauses*, such that one party agrees to compensate the other for specific types of losses or damages. Indemnification and contractual liability played significant roles in the establishment of the nuclear power industry in the United States, under a federal indemnity scheme established by the Price-Anderson Act.[37] Internationally, the Vienna Convention on Civil Liability for Nuclear Damage created a framework combining private liability, state guarantees, and pooled industry resources.[38] The use of such contractual arrangements in an AI context is still evolving.[39]

[32] S. Burton et al. *Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective*, Feb 2020
[33] H. Zech, *Liability for AI: Public Policy Considerations*, Jan. 2021
[34] Y. Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 2017
[35] S. Chesterman, *Artificial intelligence and the limits of legal personality*, 2020
[36] T. Sourdin, *Judge v Robot?: Artificial intelligence and judicial decision-making*, 2018
[37] M. Kovac, *Autonomous Artificial Intelligence and Uncontemplated Hazards: Towards the Optimal Regulatory Framework*, 2022
[38] R. Trager et al., *International governance of civilian AI: A jurisdictional certification approach*, Aug. 2023
[39] Hannes Claes & Maarten Herbosch, M. *Artificial Intelligence and Contractual Liability Limitations: A Natural Combination?*, 2023

2.4 The Role of AI Standards in Mitigating Liability Risks

By adhering to rigorous industry standards for ethical and safe AI development and deployment, such as those developed by IEEE[40], ISO[41], or national standards bodies,[42] companies can significantly reduce their liability risks. Firstly, by meeting such standards, companies demonstrate their commitment to due diligence and duty-of-care, which can serve as compelling evidence in liability litigation. As courts grapple with the complexities of AI-related harms, standard-setting bodies are likely to serve as guideposts for determining what constitutes reasonable care. Even more critically, by taking meaningful steps to adhere to rigorous standards regarding transparency, accountability, and algorithmic bias, companies will actively mitigate potential harms. Yet the leverage provided by liability will be critical: robust liability regimes will create powerful incentive structures, encouraging companies to adopt and implement industry standards, while providing a framework for accountability when those standards are not met.[43]

3. A global view: existing AI liability policy across jurisdictions

This chapter presents an overview of AI liability developments in different countries and regions, assembled by our international team to assess the state-of-play regarding AI liability. Analyzing the information on developments and initiatives in different parts of the world our team was able to find, it is clear that this critical governance conversation is most developed in the European Union, while the majority of countries and regions globally are in the early stages in their attention to the topic.

Africa

To date, the African Union (AU) and African nations have not focused on frameworks for addressing liability for harms related to AI and digital technologies. The African Union AI Strategy[44] and the AU Digital Compact[45] emphasize the need for accountability to protect consumers and promote ethical AI practices. They encourage member states to contemplate the ethical ramifications and legal obligations of AI technologies but neither framework highlights liability as a governance tool to account for the impact of AI on consumers. The legal structures governing consumer and product liability differ

[40] IEEE Standards Association, *The Ethics Certification Program for Autonomous Intelligent Systems (ECPAIS)*, Retrieved September, 2024
[41] ISO, *The International Organization for Standardization, ICO/IEC 42001: 2023, Information Technology - Artificial Intelligence Management System*, 2023
[42] NIST, *Artificial Intelligence Risk Management Profile: Generative Artificial Intelligence Profile*, July 2024
[43] C. Frattone, *Reasonable AI and Other Creatures. What Role for AI Standards in Liability Litigation?*, 2022
[44] African Union, *Continental Artificial Intelligence Strategy*, July 2024
[45] African Union, African Digital Compact, August, 2024

significantly from one country to another with some nations having consumer protection laws that encompass elements of product liability, and others significantly lacking in these protections.

The African Union's Continental AI Strategy[46] highlights and emphasizes the importance of ensuring responsible AI use, particularly when addressing fairness and accountability in decision-making. The Strategy calls for regulatory frameworks that can address biases, ensure inclusivity, and hold the right stakeholders accountable—whether that be developers, service providers, or financial institutions. However, while the AU highlights the importance of consumer protection in AI, there is a critical accountability gap, as no framework has been established yet.

Meanwhile, AI-related risks are growing rapidly. AI-driven lending algorithms in some countries promise greater financial inclusion, yet may inadvertently exacerbate inequalities as they are built on biased, historical data.[47] As the training data for the lending algorithm heavily reflects urban, male users who are more digitally active, people from rural areas with limited digital footprints or less access to mobile technology may be deemed less creditworthy, even if they have a history of responsible financial behavior. This kind of bias can deepen financial exclusion and perpetuate inequalities, as marginalized groups, including women and members of rural communities, may be less likely to receive loans or other financial services. Questions of liability arise regarding who is responsible for correcting these errors—the AI developer, the service provider, or the financial institutions. Deepfakes are increasingly prevalent in some African countries and present an additional category of AI-promoted risk. Without any mechanism for recourse or accountability, these threaten not only defamation but have the potential to incite social unrest and disrupt political stability.

Thus, the need for harmonization of existing legal approaches across the African continent remains an urgent priority. The primary challenge in developing an effective liability framework is achieving uniformity in laws across different jurisdictions, alongside establishing robust mechanisms for implementation, compliance, and enforcement.

Europe

The European Union has invested considerable study in confronting the complexities related to AI liability as a lever of AI governance. By revising its *Product Liability Directive*

[46] African Union, *Continental Artificial Intelligence Strategy*, July 2024
[47] B. E. Abikoye & C. Agorbia-Atta, *How artificial intelligence and machine learning are transforming credit risk prediction in the financial sector*, 2024

(PLD), the EU has explicitly begun to address harms caused by AI software. PLD for example lowers the burden-of-proof, to allow for redress for harms created by opaque and autonomous AI systems[48]. By clarifying that software falls within the scope of 'product', and extending liability to cases of cybersecurity vulnerabilities, the revised PLD creates a more comprehensive framework for addressing AI-related harms. This approach not only incentivizes developers and deployers to adhere to consistent standards but also prevents regulatory arbitrage across different regions. Importantly, PLD alleviates the burden of proof for victims and extends compensable damage to include psychological harm and data loss and through that makes it easier for affected individuals to seek redress.[49]

EU's standalone *AI Liability Directive* (AILD) has lingered in the proposal stage. A recent study by the European Parliamentary Research Service suggested that the AILD should be broadened to encompass a more comprehensive software liability framework, "to prevent market fragmentation and enhance clarity across the EU".[50] The study recommended a mixed framework: for AI systems that have been legally banned under the AI Act, strict liability should be assumed for damages caused; elsewhere, the strict liability standard was recommended for high-risk AI systems causing "illegitimate" harms[51].The EPRS recommended expanding the scope of the AILD so it covers not only "high-risk" but "high-impact" AI systems to encompass general purpose AI, autonomous vehicles, and other applications not classified as high-risk under the AI Act. The study calls for more explicit liability coverage for AI discrimination cases; closer attention to liability for built-in biases, privacy and intellectual property violations in general purpose AI systems, and greater harmonization of definitions with the already ratified AI Act.

Assigning liability across the AI value chain is challenging. The EPRS endorses further study of three policy options: 1) Presumption of an equal share of liability; 2) Exempting or protecting small and medium enterprises (SMEs) from the more rigorous liability expectations; 3) Protecting downstream parties such that upstream actors (particularly those with highly dominant market positions) are deemed more responsible for harms and for providing financial recourse.[52]

[48] European Parliament, *New Product Liability Directive - Q4 2020*. Sept, 2024
[49] Ibid.
[50] European Parliamentary Research Service, *Proposal for a directive on adapting non-contractual civil liability rules to Artificial Intelligence*, Sept. 2024
[51] As distinguished from "legitimate harm" models which might result in an individual rightfully being excluded from an award or benefit
[52] European Parliamentary Research Service, *Proposal for a directive on adapting non-contractual civil liability rules to Artificial Intelligence*, Sept. 2024

India

While India has not yet enacted AI-specific regulations, existing legal frameworks provide some basis for addressing AI-related liability. The Information Technology Act, 2000, which establishes liability for content on websites, could potentially extend to AI service providers – holding them responsible for content available through their platforms. Additionally, the Digital Personal Data Protection Act, 2023 introduces liability for the misuse of personal data, which could apply to AI systems processing such information. While not explicitly targeting AI, these laws create a framework where AI developers and deployers could be held liable for harmful or unlawful outcomes in areas of content moderation and data protection.

Iran (Islamic Republic of)

There is no law in Iran addressing liability for AI directly but according to the *Civil Liability Law*,[53] anyone who – without legal authorization – intentionally or negligently causes harm to another person's life, health, property, freedom, reputation, commercial reputation, or any other right established by law, resulting in material or moral damage, is responsible for compensating the damage caused. In cases where the harmful act has caused material or moral damage to the injured party, the court, after investigation and proof of the matter, will order the perpetrator to compensate for the damages. If the harmful act has caused only one type of damage, the court will order the perpetrator to compensate for that specific type of damage. So, "anyone" in this law could be interpreted to "any AI machines".

According to the National Policy of AI of I.R. Iran,[54] a set of ethical principles will guide the responsible and value-based development and use of AI technology, based on Islamic values. These principles are observed by professionals and others involved in the design, production, and utilization of AI, creating mutual rights. Examples of AI ethical issues include: respecting privacy, upholding individual and social rights, ensuring social security, fairness, explainability, transparency, non-discrimination and bias, accountability, alignment with the values and norms of Islamic society, responsibility, trust, and preventing misuse of technology. The goal of AI ethics is to optimize the beneficial impact of AI on society and human life while reducing the risks and unintended consequences of its use, based on Islamic values and beliefs. Article 5 of the national policy underscores attention to justice, dignity, rights, and the physical, mental, and

[53] Islamic Republic of Iran, *Civil Code of the Islamic Republic of Iran*, amended in 1982
[54] Center for AI and Digital Policy (CAIDP), *Artificial Intelligence and Democratic Values 2023: Iran* , April, 2024

psychological well-being of individuals in the mechanisms of training and utilizing artificial intelligence.[55]

China

In July 2023, the Cyberspace Administration of China (CAC) along with six other Chinese regulators, jointly issued Interim Measures for the Management of Generative AI Services[56] reflecting feedback from different stakeholders on previously released draft measures, and setting out the rights and responsibilities of providers and users of AI. Together, these measures establish compliance requirements for generative AI service providers, including obligations related to data sourcing, intellectual property rights, personal information protection, and content accuracy. Service providers must ensure the legitimacy of their data sources, obtain consent for using personal information, and take measures to improve training data quality. The framework also mandates labeling of AI-generated content, particularly for "deep synthesis" services, and requires providers to prevent the generation or transmission of illegal content. Violations of these obligations can result in administrative or criminal penalties, effectively creating a form of administrative liability for AI service providers.[57]

Hong Kong SAR authorities have actively sought changes to update copyright law to bolster AI development in an effort to keep pace with AI developments as the city aims to become a regional IP trading center.[58] The bureau added that it has reviewed the relevant legislation in Hong Kong and other jurisdictions, as well as the prevailing market situation.

Bangladesh

Bangladesh faces a stark digital divide, with a significant percentage of the population lacking access to the internet.[59] The government data from Bangladesh Sample Vital Statistics shows that the prevalence of internet usage among the rural population is around 37 percent and it is around 54 percent among urban population, implying a gap of 17 percent. Similarly, it finds that such a gap also persists between males and females by around 13 percent. This stark digital divide has far-reaching implications for

[55] Tehran Times, *Govt starts implementing national document on AI development*, July 24, 2024
[56] PwC: Tiang and Partners, Regulatory and legislation: China's Interim Measure for the Management of Gen AI Services, August, 2023
[57] Ibid.
[58] The Government of Hong Kong S.A. Region of China Intellectual Property Department, *Public Consultation on Copyright and Artificial Intelligence*, July 2024
[59] Khawaja Sazzad Ali & Anisur R. Faroque, *Addressing the Complexity of the Digital Divide and the Role of Government in Addressing It: Role of Government in Bridging the Digital Divide*, 2023.

Bangladesh's development, limiting access to information, education, and economic opportunities, and exacerbating existing profound inequalities. Addressing the digital divide is crucial for Bangladesh's progress and its ability to harness the potential of AI. The lack of internet connectivity also poses challenges for enforcing AI-related regulations, as it can make it difficult to monitor and regulate AI activities in remote areas.

Any AI liability framework in Bangladesh must be coupled with initiatives to bridge the digital divide, including investments in digital infrastructure, promotion of digital literacy, and efforts to make internet access more affordable and widespread.

Indonesia

In 2020, Indonesia reached a milestone in formally recognizing AI as a distinct business sector[60] via *The Indonesian National Strategy on Artificial Intelligence*.[61] The strategy designated the Ministry of Communication and Informatics to formulate ethical guidelines for AI. While further regulations are anticipated,[62] Indonesia has not yet established specific regulations overseeing AI. However, several existing legal frameworks can be leveraged for this purpose, including the Personal Data Protection Bill.[63] The liability mechanisms under this framework are divided into two parts: first is a criminal accountability mechanism, which applies solely to individuals deliberately engaging in acts intended to breach and/or misuse personal data. Breaches resulting, instead, from negligence are subject exclusively to administrative sanctions, with ambiguous remedy mechanisms. To date, this law does not comprehensively address accountability measures for potential violations and abuses carried out by the State.

Japan

In February, 2024, the ruling party of Japan issued an AI white paper which proposed an AI Basic Law in February 2024 promoting AI safety[64]; however, the proposal stresses voluntary measures (soft law), applying hard law to extreme risks presented by high-risk AI. It proposes the establishment of an AI Safety Institute (AISI) as being key to addressing harms by AI. The AISI will undertake the following measures: investigations,

[60] This recognition was actualised by introducing Ministerial Regulation No. 3/2021 issuance by the Ministry of Communication and Information Technology, designating it as the governing body for emerging technologies such as AI, Blockchain, and IoT. The implementation of Government Regulation No. 5/2021 further solidified this recognition.
[61] Stranas AI, *Indonesian National Strategy on Artificial Intelligence, 2020.*
[62] Government Regulation Number 71 the Year 2019, regarding implementing Electronic Systems and Transactions, or cloud computing and its procurement regulation
[63]The PDP framework encompasses notice and consent mechanisms, the right to be forgotten, transparency and documentation requirements, and emphasizes special considerations for children and individuals with disabilities.The PDP is complemented by consumer protection law and human rights law to address any privacy or human rights breaches.
[64]LDP Japan, *AI White Paper 2024,* April, 2024

standards, creation, developing human talent to address AI safety, fostering third-party certifications and international harmonization. The proposed policy is aimed at promoting Japan as the "world's most AI-friendly country." The white paper makes no reference to liability laws or frameworks.

South Korea

The South Korean legislature proposed a bill focused on AI liability in February, 2023.[65] The proposed law would hold high-risk AI business operators liable for damages caused to users when they are in violation of the obligations of the Act. The obligations include risk assessments, user notifications, human oversight for both developers and deployers. The bill includes exemptions for defects causing harms which could not be anticipated given the current state of science. The law would establish an "Artificial Intelligence Dispute Mediation Committee" to handle liability disputes and compensation claims. The bill also promotes insurance coverage for high-risk AI businesses to balance accountability with support for emerging technologies.

Singapore

In 2020 and 2021, the Singapore Academy of Law issued two reports on AI liability: "Report on the Attribution of Civil Liability for Accidents Involving Autonomous Cars"; and "Report on Criminal Liability, Robotics and AI Systems."[66] These propose that for intentional AI harms, existing laws could be amended to be fit-for-purpose. For civil harms that are non-intentional, the report notes several potential approaches: framing AI systems as legal personalities (such as with corporations or nation-states); creating a new category of legal offense for computer programs that commit harms; applying workplace safety legislation as a model – imposing liability on specified entities as determined within a chain of responsibility.

Brazil

Brazil's draft bill on AI governance, PL 2338/2023, establishes a clear framework for civil liability related to AI systems.[67] For high-risk or excessive risk AI systems, the bill specifies that suppliers or operators are strictly liable for damages caused, to the extent of their participation in the damage, regardless of the system's degree of autonomy. For AI systems not classified as high-risk, the bill establishes a presumption of fault, with the burden of proof shifted in favor of the victim. There are exemptions through which AI actors may not be held liable, such as when they can prove they did not deploy the AI

[65] Korea, *Bill on Artificial Intelligence Liability*, Feb., 2023
[66] Singapore, *Report Series: The Impact of Robotics and Artificial Intelligence on the Law*, 2021
[67] Brazil, Bill 2338/2023, 2023

system or when damage results exclusively from victim or third-party action. These provisions create a comprehensive liability framework across the AI lifecycle, with stricter standards for high-risk systems and maintained protections for consumers.

United States

Liability laws in the U.S. have not been updated to date to address harms from AI and algorithmic systems. However, there has been increased discussion in the U.S. tech policy world regarding the potential for liability laws to play a critical role in AI governance. An influential US tech policy think tank, Center for Humane Technology, recently published a proposal which placed liability at the center of its legislative efforts for many of the same reasons described in this discussion paper.[68] The authors note, "Current legal precedent does not define the status of AI with respect to product liability law….Liability would provide a framework for protection and legal recourse to address immediate and emerging harms from unregulated, highly powerful AI systems, especially as capabilities increase and use proliferates."

Proposals include assigning a "duty of care" to AI developers and deployers, establishing a legal obligation to prioritize safety and harm prevention in their product design and deployment. Such efforts could integrate with independently-established and ratified standards, for example through IEEE Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)[69], such that liability risks are mitigated by careful compliance.

A proposal on AI liability reform from the Center for Urban Policy Research at Rutgers University[70] makes the case for this mechanism as a means to leverage market forces and familiar legal mechanisms in the interest of safer, more ethical AI outcomes, arguing that expanding liability laws to take algorithmic and autonomously advancing harms into account will incentivize companies to integrate safeguards in their design and deployment. The Center advocates for clearer legal standards and enforcement mechanisms, including the ability for regulatory agencies to bring liability complaints against developers for negligence.

[68] Center for Humane Technology. (2024). *A Framework for Incentivizing Responsible Artificial Intelligence Development and Use*. Retrieved September 24, 2024
[69] IEEE Standards Association, *The Ethics Certification Program for Autonomous Intelligent Systems (ECPAIS)*, Retrieved September, 2024
[70] M.H. Pfeiffer, *First Do No Harm: Algorithms, AI, and Digital Product Liability,* Sept. 2023

Australia

The Australian Government is in the midst of a public, comprehensive consultation in the effort to provide effective governance, and "best practice for safety".[71] While Australia's current AI Safety Standards are voluntary, the Department of Industry, Science and Resources has developed a draft document on "mandatory guardrails for AI in high-risk settings"[72] which notes, "some AI characteristics are limiting the ability of existing laws to effectively prevent or mitigate risks….Examples include: clarifying accountability and ensuring legal responsibility is distributed appropriately to developers and deployers best placed to manage the causes of potential harms from AI decisions and applications, particularly as many existing laws were originally drafted on the presumption that humans are taking actions and making decisions."

Extrapolating from another Australian industry, the global AI governance community might take inspiration to help in assigning liability for AI harms from Australia's "chain of responsibility" model within its Heavy Vehicle National Law.[73] Under this model, each party in the value chain is responsible for ensuring that the next party can meet established safety and quality standards. The clarity of this framework might help to lighten some of the complexities surrounding AI harms, establishing a duty of care such that each entity in the AI system's lifecycle takes responsibility for verifying the capabilities and standards of the next.

4. 4. Conclusion and Recommendations:

Through this discussion paper, we hope to have amplified the conversation about a notable gap in global AI governance – applying liability frameworks as an indispensable lever to incentivize safe and ethical outcomes, and to offer recourse for harms. While researchers and policymakers around the world have acknowledged the need to clarify the legal complexities that accompany AI liability, this conversation has thus far been most prominent in the European Union. A disparity in liability frameworks will create highly risky vulnerabilities, particularly for individuals and communities in the Global South where AI-related harms, from biased lending algorithms to unchecked deep fakes, are already pervasive.

[71] Australian Government, *Promoting safe and responsible AI,* retrieved Sept. 2024
[72] Australian Government, *Introducing mandatory guardrails for AI in high-risk settings*, *Sept 2024*
[73] B. Walker-Munro & Z. Assaad, Z, The Guilty (Silicon) Mind: Blameworthiness and Liability in Human-Machine Teaming, Oct. 2022

To address this gap in governance, we propose four immediate actions:

1. Establish a **Global AI Liability Task Force**, bringing together experts from diverse jurisdictions to develop harmonized principles that can be adapted across frameworks.

2. Formalize **adherence to ethical AI industry standards** – such as those by IEEE and ISO – into liability frameworks – to incentivize AI companies to rigorously implement the voluntary safeguards delineated by standards-setting bodies.

3. Investigate the potential applicability of **a "chain of responsibility" framework for AI liability** to clarify accountability across the complex AI lifecycle.

4. Develop **capacity-building initiatives focused on AI liability** in Global Majority countries, addressing both the technical and legal aspects of enforcement, coupled with initiatives to bridge this divide, including investments in digital infrastructure and promotion of digital literacy, to ensure effective implementation and enforcement of AI regulations.

By elevating the conversation on AI liability as a lever of governance among global policy stakeholders, we hope to help level the playing field for AI, mitigate risks, establish a fair means for obtaining recourse for harms, and increase the likelihood that the benefits of AI are realized responsibly and equitably around the world.

Authors

Policy Network on AI Sub-group on Liability as a mechanism for supporting AI accountability

Team leaders

Caroline Friedman Levy and Umut Pajaro Velasquez

Penholders

Jasmine Ko (DotAsia, Hong Kong China); Lufuno T Tshikalange, Orizur Consulting Enterprise, Republic of South Africa; Omor Faruque (Dynamic Teen Coalition & Project OMNA, Bangladesh); Sameer Gahlot (National Internet Exchange of India, India); Aji Fama Jobe(Women Techmakers Banjul & ISOC Youth Standing Group, The Gambia); Antara Vats (ARTICLE19, India)

About the Policy Network on Artificial Intelligence

The Policy Network on Artificial Intelligence (PNAI) addresses policy matters related to artificial intelligence and data governance. It is a global multistakeholder effort hosted by the United Nations' Internet Governance Forum, providing a platform for stakeholders and changemakers in the AI field to contribute their expertise, insights, and recommendations. PNAI's primary goal is to foster dialogue and contribute to the global AI policy discourse. Participation in and contribution are open to everyone.