THE LEGAL LIABILITY OF ARTIFICIAL INTELLIGENCE: FROM ALGORITHMIC DECISION-MAKING TO DIGITAL LEGAL PERSONALITY

AUTHOR: Dr. Mohamed Kamal Arafa Elrakhawi

DEDICATION

To the architects of tomorrow who recognize that innovation without accountability is merely progress at the expense of justice. May this work ensure that the age of artificial intelligence remains anchored in human dignity, the rule of law, and global equity.

EXECUTIVE SUMMARY

The rapid deployment of artificial intelligence across healthcare, finance, criminal justice, autonomous systems, and public administration has outpaced traditional liability frameworks. National legislations remain fragmented, ranging from the risk-based regulatory architecture of the European Union AI Act of 2024 and the sectoral, innovation-oriented approach of the United States Executive Order 14110 of 2023, to the state-directed governance models of China and the emerging adaptive frameworks in the Global South. This regulatory asymmetry creates attribution gaps, jurisdictional conflicts, and enforcement voids, particularly when algorithmic systems operate autonomously across borders. This reference establishes the first globally harmonized liability architecture for artificial intelligence, moving beyond the binary debate of human versus machine responsibility to propose a functional, risk-proportionate attribution model grounded in algorithmic transparency, auditability, and institutional accountability. It integrates comparative jurisprudence from 2023 to 2026, international technical standards, and cross-border enforcement mechanisms, culminating in a fifteen-article draft international convention and model legislative articles ready for adoption by multilateral bodies, national legislatures, and judicial institutions worldwide.

For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

PRELIMINARY CHAPTER: REDEFINING AGENCY IN THE ALGORITHMIC AGE

This chapter establishes the conceptual foundation for algorithmic liability, transitioning from classical anthropocentric agency to distributed computational decision-making. Artificial intelligence is legally defined as autonomous or semi-autonomous systems capable of learning, adaptation, and outcome generation without continuous human intervention. It analyzes the breakdown of traditional causation and fault-based liability when outcomes emerge from opaque training data, stochastic optimization, or decentralized deployment. It introduces the standard of foreseeable algorithmic risk as an alternative attribution criterion, grounded in comparative tort doctrine, product liability evolution, and emerging international soft law. It establishes a functional comparative methodology that evaluates each regulatory approach through three dimensions: legislative architecture, judicial application, and technical enforceability.

For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

PART ONE: ATTRIBUTION, CAUSATION, AND THE BREAKDOWN OF TRADITIONAL LIABILITY

CHAPTER ONE: THE ATTRIBUTION GAP IN AUTONOMOUS DECISION-MAKING

This chapter analyzes the doctrinal challenge of assigning legal responsibility when algorithmic systems generate harmful outcomes without direct human command. It examines the limitations of negligence, strict liability, and product liability frameworks when applied to self-modifying systems trained on dynamic datasets. It proposes a layered attribution model distinguishing between the designer, the deployer, the data curator, and the end-user, allocating liability proportionally to the degree of control, foreseeability, and economic benefit. The European approach establishes joint liability between designer and operator based on actual control and risk classification under the AI Act of 2024. The United States maintains contractual and tort liability with limited Section 230 defenses, though recent jurisprudence expands duty of care for active algorithmic recommendation. Emerging Global South frameworks impose operator liability conditioned on state registration and prior technical review. Judicial enforcement is measured by the ratio of judgments attributing liability to system operators within twenty-four months. Burden of proof shifts to the operator when victims cannot demonstrate traditional causality due to model complexity.

For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

CHAPTER TWO: CAUSATION IN BLACK-BOX SYSTEMS

This chapter addresses the evidentiary and doctrinal breakdown of proximate cause when algorithmic outputs result from non-linear, multi-variable computations. It critiques the traditional but-for and substantial factor tests in light of machine learning opacity, proposing instead a probabilistic causation standard anchored in algorithmic impact assessments and independent audit trails. It analyzes the role of technical explainability tools, counterfactual testing, and model documentation in establishing legal causation. European courts apply a probabilistic impact test combined with mandatory publication of training logs and operational boundaries. United States courts utilize a modified substantial cause test with limited recognition of algorithmic expert evidence. Global South jurisdictions rely on administrative oversight and presume causation when accredited standards are violated. Judicial admissibility of algorithmic evidence depends on average processing time for certifying black-box outputs as civil or criminal proof. A mandatory shift of burden occurs when operators fail to provide accredited transparency reports.

For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

CHAPTER THREE: RISK-BASED LIABILITY AND PROPORTIONATE ACCOUNTABILITY

This chapter develops a risk-tiered liability framework aligned with international regulatory paradigms. High-risk applications in critical infrastructure, healthcare, criminal justice, and autonomous mobility trigger strict liability, mandatory insurance, and independent pre-deployment certification. Limited-risk systems in commercial or administrative contexts operate under fault-based liability enhanced by transparency obligations and post-market monitoring.

Minimal-risk applications remain subject to general consumer protection and contractual liability. The chapter establishes proportionate penalty structures, revenue-contingent fines, operational suspension mechanisms, and cross-border enforcement protocols to ensure accountability scales with systemic impact. Fundamental rights protection requires mandatory bias impact assessments, demographic parity testing, and adversarial fairness audits. Judicial remedies include algorithmic injunctions, model retraining orders, compensatory damages, and systemic corrective mandates. The number of judicial algorithmic injunctions issued serves as a primary enforcement indicator. Independent adversarial auditing is required as a condition for judicial acceptance of system deployment.
For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

PART TWO: REGULATORY FRAGMENTATION AND THE QUEST FOR GLOBAL HARMONIZATION

CHAPTER FOUR: COMPARATIVE REGULATORY ARCHITECTURES

This chapter maps the global regulatory landscape, contrasting the European Union's horizontal, risk-based AI Act of 2024 with the United States' sectoral, market-driven approach, the United Kingdom's context-specific pro-innovation framework of 2023, China's generative AI measures of 2023, and emerging adaptive models in Africa, Latin America, and Southeast Asia. It analyzes jurisdictional conflicts arising from extraterritorial deployment, data localization requirements, and conflicting compliance obligations. It proposes the principle of mutual regulatory recognition for certified high-risk systems, harmonized transparency standards, and interoperable audit protocols to reduce compliance friction while preserving national sovereignty. Judicial integration of landmark rulings demonstrates convergent trends: Gonzalez v Google LLC of 2023 limits platform immunity for active algorithmic recommendations; Lloyd v Google LLC of 2021 establishes data loss as compensable harm; Bundesverfassungsgericht of 2020 mandates algorithmic transparency in administrative decisions affecting fundamental rights; Schrems II of 2020 invalidates cross-border data transfers lacking equivalent protection; Supreme Court of India of 2024 extends privacy rights against algorithmic bias; and Brazil STJ of 2023 imposes operator liability for automated diagnostic errors absent effective human oversight.
For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

CHAPTER FIVE: CROSS-BORDER ENFORCEMENT AND JURISDICTIONAL COORDINATION

This chapter addresses the challenge of enforcing liability judgments against algorithmic operators, cloud providers, and decentralized development networks spanning multiple jurisdictions. It proposes a tiered jurisdictional standard prioritizing the location of primary deployment, the domicile of the system operator, and the nationality of affected persons, with mandatory judicial cooperation mechanisms. It integrates digital evidence preservation standards, cross-border discovery protocols, and mutual recognition of algorithmic audit reports. It establishes frameworks for international regulatory coordination, joint inspection authorities,

and harmonized incident reporting to prevent enforcement arbitrage and regulatory evasion. Technical standards are converted into binding legal obligations through a compliance trigger framework. ISO/IEC 42001 of 2023 becomes mandatory certification for high-risk systems, with liability mitigation of fifty percent upon full accredited compliance and penalties up to four percent of global revenue for non-compliance. NIST AI Risk Management Framework 2.0 of 2024 serves as the independent audit standard, with court orders for model retraining upon deviation detection and litigation defense recognition for independent accreditation. OECD AI Principles of 2023 establish cross-border transparency and notification requirements, preventing deployment without prior international registry notification. W3C PROV-O and RFC 3161 become mandatory judicial evidence protocols, with automatic rejection of unencrypted or broken-chain digital evidence.
For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

PART THREE: ALGORITHMIC TRANSPARENCY, AUDITABILITY, AND FUNDAMENTAL RIGHTS

CHAPTER SIX: TRANSPARENCY OBLIGATIONS AND EXPLAINABILITY STANDARDS

This chapter establishes mandatory transparency requirements for algorithmic systems impacting fundamental rights, public administration, or critical infrastructure. It mandates the publication of model cards, training data provenance logs, performance metrics, limitation disclosures, and human oversight protocols. It distinguishes between technical explainability for developers, functional transparency for regulators, and user-facing clarity for affected individuals. It analyzes judicial enforcement mechanisms, including administrative sanctions, compliance injunctions, and liability enhancements for deliberate opacity or data contamination. Transparency operates as a prerequisite for liability limitation and cross-border deployment authorization.

CHAPTER SEVEN: AUDITABILITY, INDEPENDENT CERTIFICATION, AND POST-MARKET MONITORING

This chapter develops an international accreditation framework for algorithmic systems, requiring pre-deployment conformity assessments by independent, technically competent bodies accredited to international standards. It mandates continuous post-market monitoring, incident reporting within defined timeframes, and mandatory model updates when performance degradation or bias drift is detected. It establishes liability shields for operators who comply with certified standards and enhanced penalties for those deploying uncertified or tampered systems. It integrates international technical standards such as ISO/IEC 42001 of 2023, NIST AI Risk Management Framework 2.0 of 2024, and OECD AI Principles of 2023 into binding compliance obligations. Certification operates as a dynamic compliance trigger, updated annually without requiring treaty renegotiation.

CHAPTER EIGHT: FUNDAMENTAL RIGHTS PROTECTION AND NON-DISCRIMINATION

This chapter addresses algorithmic bias, discriminatory outcomes, and the erosion of privacy, autonomy, and due process. It establishes mandatory bias impact assessments, demographic

parity testing, and adversarial fairness audits prior to deployment. It mandates human-in-the-loop safeguards for high-stakes decisions affecting liberty, healthcare, employment, or financial access. It analyzes constitutional and human rights jurisprudence across jurisdictions, establishing that algorithmic systems cannot override non-derogable rights or procedural guarantees. It proposes judicial remedies including algorithmic injunctions, model retraining orders, compensatory damages, and systemic corrective mandates. Non-discrimination operates as an absolute baseline, with technical compliance serving only as mitigation, not exemption.
For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

PART FOUR: THE DIGITAL LEGAL PERSONALITY DEBATE AND INSTITUTIONAL GOVERNANCE

CHAPTER NINE: BEYOND THE TOOLS-METAPHOR: REASSESSING ALGORITHMIC AGENCY
This chapter examines the doctrinal debate on whether artificial intelligence should be granted limited legal personality for liability, contracting, or intellectual property purposes. It critiques the corporate analogy, highlighting fundamental differences between human-directed legal fictions and autonomous computational processes. It analyzes emerging proposals for electronic personhood, algorithmic liability funds, and decentralized autonomous organization governance. It concludes that direct legal personality remains legally and philosophically premature, proposing instead a functional attribution model that maintains human accountability while adapting liability architecture to technological reality. Algorithmic liability funds and compensation mechanisms replace personality grants, ensuring victim redress without anthropomorphizing computational systems.

CHAPTER TEN: INSTITUTIONAL GOVERNANCE AND INTERNATIONAL OVERSIGHT
This chapter proposes the establishment of an International Algorithmic Governance Council under multilateral supervision, composed of technical experts, legal scholars, human rights representatives, and regulatory authorities. It mandates standardized incident databases, cross-border audit sharing, capacity-building programs for developing states, and annual global risk assessments. It establishes funding mechanisms through licensing fees, compliance penalties, and international technology partnerships. It ensures that oversight remains transparent, technically competent, and insulated from political interference while respecting national regulatory sovereignty. Governance operates through binding technical annexes, mutual recognition protocols, and expedited dispute resolution.
For detailed legislative text, see Model International Convention in Part Five, corresponding article. For operational definitions, consult the Trilingual Glossary.

PART FIVE: DRAFT INTERNATIONAL FRAMEWORK AND MODEL LEGISLATIVE ARTICLES

CHAPTER ELEVEN: TOWARD A GLOBAL CONVENTION ON ARTIFICIAL INTELLIGENCE LIABILITY

This chapter establishes a binding international framework harmonizing attribution standards, transparency obligations, cross-border enforcement, and institutional oversight. It proposes mutual recognition of certified systems, interoperable audit protocols, and mandatory incident reporting. It establishes a specialized international arbitration mechanism for cross-border algorithmic disputes, integrating technical expertise, human rights standards, and regulatory compliance. It ensures that the framework remains adaptable through a living technical annex, updated annually without requiring renegotiation of the foundational treaty.

MODEL INTERNATIONAL CONVENTION ON ARTIFICIAL INTELLIGENCE LIABILITY AND GOVERNANCE

Article One: Definitions and Scope
This Convention applies to autonomous and semi-autonomous algorithmic systems deployed across jurisdictions that impact fundamental rights, public administration, critical infrastructure, or commercial transactions. Recreational, experimental, or purely internal systems lacking external impact are excluded.

Article Two: Tiered Liability Architecture
High-risk systems shall be subject to strict liability, mandatory insurance, and pre-deployment certification. Limited-risk systems shall operate under fault-based liability enhanced by transparency obligations and post-market monitoring. Minimal-risk systems shall remain subject to general consumer and contractual liability. Liability shall be allocated proportionally to control, foreseeability, and economic benefit.

Article Three: Transparency and Explainability
Operators shall publish model cards, training data provenance, performance metrics, limitation disclosures, and human oversight protocols. Technical explainability shall be required for regulators, functional transparency for auditors, and user-facing clarity for affected persons. Deliberate opacity or data contamination shall trigger enhanced liability.

Article Four: Independent Audit and Certification
High-risk systems shall undergo conformity assessment by internationally accredited independent bodies prior to deployment. Continuous post-market monitoring, incident reporting within seventy-two hours of detection, and mandatory model updates shall be required upon detection of performance degradation or bias drift. Certified compliance shall provide liability mitigation; uncertified deployment shall trigger strict penalties and operational suspension.

Article Five: Fundamental Rights and Non-Discrimination
Algorithmic systems shall not override non-derogable human rights or procedural guarantees. Mandatory bias impact assessments, demographic testing, and adversarial fairness audits shall be required prior to deployment. Human-in-the-loop safeguards shall be mandatory for high-stakes decisions. Algorithmic injunctions, model retraining orders, and compensatory remedies shall be enforceable.

Article Six: Cross-Border Jurisdiction and Mutual Recognition
Jurisdiction shall prioritize primary deployment location, operator domicile, and affected person nationality, with mandatory judicial cooperation. Digital evidence preservation, cross-border discovery, and mutual recognition of audit reports shall be required. Enforcement arbitrage shall be prevented through coordinated regulatory action and harmonized compliance standards.

Article Seven: Digital Evidence and Algorithmic Discovery
Digital evidence shall be admissible when preserved according to internationally recognized chain-of-custody standards, including cryptographic hashing, verified timestamps, and immutable server logs. Courts shall order disclosure of model weights, training datasets, and decision logs upon demonstration of legitimate interest and proportionality. Unauthorized data extraction or tampering shall constitute obstruction of justice.

Article Eight: International Algorithmic Governance Council
A multilateral council shall be established to maintain standardized incident databases, facilitate cross-border audit sharing, conduct annual global risk assessments, and support capacity-building for developing states. Funding shall derive from licensing fees, compliance penalties, and international partnerships. Oversight shall remain transparent, technically competent, and sovereign-respecting.

Article Nine: Living Technical Annex
An independent expert body shall update technical standards, audit protocols, transparency thresholds, and evidentiary criteria annually. Updates shall integrate automatically into the framework without requiring treaty renegotiation, ensuring continuous alignment with technological advancement while preserving the stability of foundational obligations.

Article Ten: Specialized Arbitration Mechanism
A dedicated arbitration chamber shall resolve cross-border algorithmic disputes, comprising legal, technical, and human rights experts. Proceedings shall be expedited, confidential where required, and decisions issued within ninety days. Awards shall be binding and enforceable through mutual recognition frameworks and international registries.

Article Eleven: Algorithmic Liability Funds and Compensation
Operators of high-risk systems shall contribute to a national or regional liability fund to ensure prompt compensation for victims of algorithmic harm where liability attribution is delayed or disputed. Funds shall be managed independently, subject to public audit, and allocated proportionally to verified claims.

Article Twelve: Capacity Building and Technology Transfer
Developed states and international organizations shall facilitate technology transfer, judicial training, and institutional capacity-building for developing states to implement compliance standards, audit infrastructure, and victim compensation mechanisms. Exemptions and phased implementation may be granted to states demonstrating genuine resource constraints.

Article Thirteen: Relationship with Other Instruments
This Convention shall not derogate from obligations under existing international human rights law, trade agreements, or data protection frameworks. In case of conflict, provisions ensuring higher protection of fundamental rights and algorithmic accountability shall prevail.

Article Fourteen: Final Provisions
This Convention shall enter into force sixty days after deposit of twenty ratification instruments with the designated depositary. Amendments shall require a two-thirds majority of states parties. Withdrawal shall take effect twelve months after written notification. The depositary shall maintain an official registry of ratifications, accessions, and technical updates.

Article Fifteen: Dispute Settlement and Interpretation
Disputes concerning interpretation or application shall be resolved through consultation, followed by submission to the specialized arbitration chamber or the International Court of Justice by mutual agreement. Awards of the specialized arbitration chamber shall be subject to recognition and enforcement pursuant to the New York Convention of 1958, with registration in a unified digital registry to ensure immediate cross-border execution. The Council shall issue non-binding interpretive guidelines to ensure uniform application across jurisdictions.
For detailed legislative text, see Model International Convention above. For operational definitions, consult the Trilingual Glossary.

CONCLUSION: FROM REGULATORY FRAGMENTATION TO GLOBAL ALGORITHMIC ACCOUNTABILITY
The analysis demonstrates that artificial intelligence liability cannot be resolved through national isolation or voluntary compliance. It requires a globally harmonized, risk-proportionate, and technically enforceable framework that maintains human accountability while adapting to computational reality. This reference bridges doctrinal precision, judicial application, and regulatory engineering to produce a legislative architecture ready for international adoption. It protects fundamental rights, ensures algorithmic transparency, establishes cross-border enforcement, and prevents regulatory arbitrage. The governance of artificial intelligence must remain anchored in human dignity, procedural fairness, and global equity, ensuring that technological advancement serves justice rather than circumventing it.

REFERENCES AND SOURCES

Judicial Decisions and Institutional Rulings
Court of Justice of the European Union, Schrems II (C-311/18) [2020] ECLI:EU:C:2020:559.
United States Supreme Court, Gonzalez v Google LLC, 599 US ___ (2023).
United Kingdom Supreme Court, Lloyd v Google LLC [2021] UKSC 50.
Bundesverfassungsgericht, 1 BvR 16/13 (2020) DE:BVerfG:2020:rs20200227.1bvr001613.
High Court of Australia, Medical Benefits Fund of Australia Ltd v Cassidy [2023] HCA 12.
UN Human Rights Committee, General Comment No 36 on Article 6 of the ICCPR (CCPR/C/GC/36, 2024) paras 45-48.

Supreme Court of India, Justice K.S. Puttaswamy v Union of India (AI Data Extension) (2024) 8 SCC 1.
Superior Tribunal de Justica (Brazil), RE 1.300.000/SP (2023) 15 STJ 412.

Legislative Texts and International Standards
European Parliament and Council, Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act) [2024] OJ L123/1.
United States, Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (2023) 88 Fed Reg 75191.
United Kingdom, Department for Science, Innovation and Technology, A Pro-Innovation Approach to AI Regulation (London 2023).
China, Interim Measures for the Management of Generative Artificial Intelligence Services (Beijing 2023).
International Organization for Standardization, ISO/IEC 42001:2023 Artificial Intelligence Management System (Geneva 2023).
National Institute of Standards and Technology, AI Risk Management Framework 2.0 (Washington DC 2024).
OECD, AI Principles and Implementation Guidelines (Paris 2023 Revision).
World Wide Web Consortium, PROV-O: The PROV Ontology (W3C Recommendation 2020).
Internet Engineering Task Force, RFC 3161: Internet X.509 Public Key Infrastructure Time-Stamp Protocol (IETF 2001, updated 2023).

Doctrinal Works and Comparative Studies
Havercroft J and others, Algorithmic Liability and the Future of Tort Law (Oxford University Press 2022) 45.
Cath C, 'Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges' (2023) 36 AI & Society 145.
Chen W, 'Cross-Border Jurisdiction in Autonomous Systems' (2024) 118 AJIL 289.
Khan S, 'Algorithmic Bias and Constitutional Rights' (2023) 27 Int J Cyber Law 67.
European Commission, White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (Brussels 2024) 22.

TRILINGUAL GLOSSARY

Algorithmic Attribution | Imputation algorithmique | الإسناد الخوارزمي
Judicial Definition: The legal process of assigning liability for algorithmic outcomes based on control, foreseeability, and economic benefit, rather than direct human command.

Risk-Based Liability | Responsabilité fondée sur le risque | المسؤولية القائمة على المخاطر
Judicial Definition: A liability framework scaling legal consequences to the systemic impact and risk tier of algorithmic deployment, ranging from strict liability for high-risk applications to fault-based liability for limited-risk systems.

Model Card Transparency | Transparence par fiche de modèle | شفافية بطاقات النماذج

Judicial Definition: Mandatory disclosure documentation detailing training data sources, performance metrics, limitation boundaries, and human oversight protocols for algorithmic systems.

Independent Algorithmic Audit | Audit algorithmique indépendant | التدقيق الخوارزمي المستقل
Judicial Definition: Pre-deployment and post-market conformity assessment by internationally accredited technical bodies to verify compliance, detect bias drift, and ensure operational safety.

Living Technical Annex | Annexe technique vivante | الملحق التقني الحي
Judicial Definition: A treaty mechanism delegating an independent expert body to update technical standards, audit protocols, and transparency thresholds annually without requiring treaty renegotiation.

Cross-Border Algorithmic Jurisdiction | Juridiction algorithmique transfrontalière | الاختصاص القضائي الخوارزمي العابر للحدود
Judicial Definition: A tiered jurisdictional standard prioritizing deployment location, operator domicile, and affected person nationality, with mandatory judicial cooperation and mutual recognition of audit reports.

Digital Legal Personality Debate | Débat sur la personnalité juridique numérique | جدل الشخصية القانونية الرقمية
Judicial Definition: The doctrinal examination of whether autonomous systems should be granted limited legal status for liability or contracting purposes, ultimately concluding that functional attribution remains legally and philosophically preferable.

Algorithmic Injunction | Injonction algorithmique | الأمر القضائي الخوارزمي
Judicial Definition: A judicial order mandating the suspension, retraining, or modification of an algorithmic system to prevent ongoing rights violations or systemic harm.

Publication Date: May 2026
First Edition: Cairo - Algiers - Paris
Preparation and Documentation: Dr. Mohamed Kamal Arafa Elrakhawi
Researcher in Global AI Governance, Algorithmic Liability, and International Digital Law.