The Responsibility Gap in Algorithmic Governance:A Legal Analysis of Criminal Liability and Due Process in Automated Administrative Decisions.

Hala Mohamed Imam Mohamed taher[1]

[1]Assistant Professor of Criminal Law - Najran University - College of Business Administration, Department of Law

Email: drhalaemam@gmail.com & hmemam@nu.edu.sa

| ORCID: /0009-0000-6929-2636

Abstract: The accelerating integration of automated decision-making (ADM) systems into public administration has generated a profound accountability deficit that existing legal frameworks are inadequately equipped to address. This paper examines the 'responsibility gap'—the structural absence of clearly assignable liability when algorithmically driven decisions produce legally harmful outcomes in administrative and judicial contexts. Drawing on a normative juridical methodology and comparative regulatory analysis, the study interrogates the doctrinal coherence of established criminal law concepts, particularly Actus Reus and Mens Rea, when applied to harms caused by autonomous computational agents. The analysis critically evaluates recent legislative instruments, including the EU Artificial Intelligence Act (2024), the General Data Protection Regulation, and Saudi Arabia's Personal Data Protection Law (PDPL), to assess the feasibility of attributing criminal and administrative liability to non-human entities. The paper advances three original contributions: a Hybrid Liability Model that distributes responsibility across the algorithmic supply chain; a Legal Risk Index (LRI) as a standardised diagnostic instrument for regulatory compliance; and a constitutional reframing of Explainable AI (XAI) as an enforceable component of procedural due process. The findings demonstrate that effective governance of algorithmic systems requires not merely technical solutions but a coherent jurisprudential architecture capable of preserving the rule of law in an increasingly automated state

Keywords: algorithmic governance; automated decision-making; criminal liability; due process; explainable AI; EU AI Act; PDPL; Legal Risk Index; responsibility gap..

Introduction

1.1 The Algorithmic Turn and the Crisis of Legal Accountability

The past decade has witnessed a decisive migration of discretionary state power from human officials to algorithmic intermediaries. Governments across the globe now deploy statistical models to adjudicate welfare eligibility, calculate criminal risk scores, allocate healthcare resources, and prioritise law enforcement interventions. This 'algorithmic turn' represents not merely a shift in administrative method, but a fundamental transformation in the constitutional relationship between the state and the citizen (Diakopoulos, 2016; Zarsky, 2016).

The consequences of this transformation are far from neutral. Complex machine learning architectures—often characterised as 'black boxes'—render the logic of state action opaque, systematically undermining the requirements of procedural fairness that underpin the rule of law. When the rationale for a consequential administrative decision cannot be articulated, traced, or contested, the foundational guarantees of due process are effectively rendered unenforceable (Wachter et al., 2017). Courts in the Netherlands, Australia, and the United Kingdom have been compelled to confront the legal implications of this opacity, with landmark rulings such as ECLI:NL:RBDHA:2020:1878 (the SyRI case) establishing that automated profiling systems may violate

the European Convention on Human Rights when deployed without adequate transparency safeguards..Yet despite this mounting jurisprudential recognition, a critical lacuna persists: existing liability frameworks remain structurally incapable of identifying a responsible human agent when an algorithmic system produces harm. This is the 'responsibility gap' that the present paper seeks to diagnose and address

1.2 Scope and Research Questions

This study is animated by three interconnected research questions. First, how must doctrinal criminal law concepts—specifically Actus Reus and Mens Rea—be reconceptualised to accommodate harm caused by autonomous algorithmic agents? Second, what liability architecture is most jurisprudentially coherent for distributing responsibility across the multi-actor algorithmic supply chain? Third, how can the constitutional principle of procedural due process be operationalised as a design-level obligation for high-risk ADM systems?

The analysis proceeds through a normative juridical methodology, employing comparative regulatory analysis of the EU AI Act (Regulation (EU) 2024/1689), GDPR (Regulation (EU) 2016/679), and Saudi Arabia's PDPL (Royal Decree M/19, 2021), supplemented by doctrinal examination of case law from multiple jurisdictions. The paper does not advance empirical claims about algorithmic performance; rather, it constructs a jurisprudential framework capable of governing such performance.

1.3 Structure of the Paper

The remainder of this paper is structured as follows. Section 2 reviews existing scholarship on algorithmic bias and the socio-legal dimensions of ADM systems. Section 3 sets out the research methodology and analytical framework. Section 4 examines the criminal law dilemma, including the doctrinal challenges to Actus Reus and Mens Rea. Section 5 analyses administrative safeguards and the evolving standard of judicial review. Section 6 presents the proposed Hybrid Liability Model, Legal Risk Index, and Fundamental Rights Impact Assessment framework. Section 7 considers sectoral case studies. Section 8 discusses the Saudi Arabia and GCC regulatory landscape. Section 9 offers conclusions and policy recommendations.

LITERATURE REVIEW

2.1 Algorithmic Bias as a Socio-Technical Phenomenon

The scholarly literature has decisively rejected the notion that algorithmic systems are inherently objective. Pasquale (2015) introduced the influential 'black box' metaphor to describe systems that convert data inputs into consequential outputs through processes deliberately obscured from public scrutiny. Subsequent empirical work, most notably Angwin et al.'s (2016) investigation of the COMPAS recidivism algorithm, demonstrated that commercially deployed risk assessment tools can exhibit systematic racial disparities—misclassifying Black defendants as high-risk at nearly twice the rate of their white counterparts.

Barocas and Selbst (2016) provided a rigorous doctrinal account of how facially neutral algorithmic systems may encode discriminatory patterns through the choice of training data, feature selection, and proxy variables. They argued that existing US anti-discrimination law is structurally ill-suited to address this form of 'disparate impact' because the law requires proof of a discrete discriminatory act traceable to a human decision-maker. This analytical insight maps directly onto the responsibility gap problem that lies at the centre of this paper.

Eubanks (2018) extended these concerns to administrative welfare systems, documenting how algorithmic tools deployed in public child services and welfare benefits administration produced systematic disadvantage for economically marginalised communities. Her work established that algorithmic bias is not merely a technical artefact but a mechanism of political economy that replicates and amplifies existing social inequities. This finding is corroborated by the Dutch SyRI litigation, in which the Hague District Court found that a government fraud-detection algorithm violated Article 8 ECHR because it disproportionately targeted low-income urban districts, effectively penalising poverty as a proxy for fraud (Rechtbank Den Haag, 2020).

2.2 The Responsibility Gap in Legal Theory

The concept of a 'responsibility gap' in relation to autonomous systems was first articulated by Matthias (2004), who argued that as machines acquire greater degrees of learning and autonomy, the connection between human agency and machine outcome becomes sufficiently attenuated to preclude conventional attribution of moral and legal responsibility. This insight has been substantially developed by subsequent scholars working at the intersection of criminal law and artificial intelligence.

Duff (2007) argued that criminal liability requires a genuine agential relationship between the defendant and the harmful outcome—a relationship predicated on the capacity for practical reasoning that autonomous systems, however sophisticated, do not possess. Floridi et al. (2018) proposed the concept of 'distributed moral responsibility' as a partial remedy, suggesting that responsibility could be allocated across the network of human agents involved in an algorithm's design, deployment, and operation, even when no single agent possesses complete knowledge of the system's behaviour. Doshi-Velez et al. (2017) argued for a technical complement to this legal solution, contending that interpretability—the capacity of a system to generate human-comprehensible explanations of its decisions—is not merely a desirable design feature but a precondition for meaningful legal accountability.

More recently, the literature has engaged with the legislative response. Wachter, Mittelstadt, and Russell (2017) provided an influential analysis of GDPR Article 22 and the contested 'right to explanation,' concluding that the Regulation's provisions are weaker than commonly assumed and do not establish a genuine obligation to provide algorithmic justifications in all cases. The EU AI Act represents a legislative attempt to address this gap directly, and several scholars have already begun to evaluate its adequacy (Veale and Borgesius, 2021; Edwards and Veale, 2022).

2.3 The GCC and Saudi Arabian Context

The emerging governance literature on algorithmic accountability in the Gulf Cooperation Council region remains comparatively sparse, a gap this paper seeks partially to address. Al-Sarihi and Mitchell (2021) noted that the digital transformation strategies of GCC states have proceeded largely in the absence of robust public law frameworks capable of constraining algorithmic power. The establishment of SDAIA in 2019 and the subsequent promulgation of the PDPL in 2021 represent a significant legislative development, though the law's provisions relating to automated decision-making remain underdeveloped compared to equivalent European instruments. Diab (2024) has argued that the PDPL's Article 25—which provides citizens with the right to object to decisions made solely on the basis of automated processing—represents a meaningful constitutional commitment, but one that requires substantial implementing regulation to achieve practical effect.

RESEARCH METHODOLOGY AND ANALYTICAL FRAMEWORK

3.1 Normative Juridical Method

This study employs the normative juridical method, which is the established standard for high-impact doctrinal legal research. This methodology involves systematic analysis of primary legal sources—legislation, judicial decisions, and constitutional instruments—evaluated against the normative standards they purport to embody, including the rule of law, procedural fairness, and the principle of equality before the law. The normative juridical approach is particularly well-suited to the present inquiry because the central question—whether existing legal categories can coherently accommodate algorithmic harm—is fundamentally a question about the internal consistency and practical adequacy of legal doctrine.

3.2 Comparative Regulatory Analysis

The paper employs a comparative methodology to evaluate how different jurisdictions have approached the governance of automated administrative decisions. The primary comparators are the European Union (EU AI Act and GDPR), Saudi Arabia (PDPL and SDAIA Ethical AI Principles), and Australia (examined through the Robodebt Royal Commission). These jurisdictions represent distinct regulatory philosophies—a precautionary rights-based model (EU), a values-driven developmental model (Saudi Arabia/GCC), and a post-hoc accountability model (Australia)—and their comparison yields generalisable insights for regulatory design.

3.3 Analytical Framework: The Four-Layer Review Model

The analytical framework applied in this paper examines ADM systems across four layers of legal scrutiny. The first layer is Legal Authority, which examines whether a valid legal basis exists for automated decision-making in the relevant domain. The second layer is Design Legality, which assesses whether the system's architecture—including its training data, objective function, and performance parameters—complies with applicable non-discrimination and data protection obligations. The third layer is Procedural Fairness, which evaluates whether the system's operation satisfies the requirements of procedural due process, including the provision of reasons and opportunities for challenge. The fourth layer is Remedial Efficacy, which examines whether adequate mechanisms exist to provide redress when algorithmic decisions cause harm.

THE LEGAL DILEMMA: RECONCEPTUALISING CRIMINAL LIABILITY FOR ALGORITHMIC HARM

4.1 Actus Reus and the Concept of Functional Acts

Traditional criminal law predicates liability on the commission of a voluntary act—Actus Reus—by a human agent possessing legal capacity. The deployment of autonomous algorithmic systems disrupts this paradigm in two important respects. First, the act that produces harm—the discriminatory output of an algorithm—may be neither directly performed nor specifically intended by any identifiable human. Second, the causation chain between human choices (in design, training, or deployment) and algorithmic outcomes may extend across complex multi-party relationships that resist reduction to the simple dyadic structure that criminal law assumes.

Scholars including Turner (2019) and Solum (1992) have proposed the concept of 'functional acts' to address this gap—the idea that certain legally relevant acts may be performed by non-human agents if those agents possess sufficient functional autonomy and a causal relationship to the harmful outcome. While intellectually innovative, this approach faces significant doctrinal resistance: criminal law's

requirement of a voluntary act is grounded in a deep commitment to the principle that punishment is only justified where an agent exercised genuine control over the prohibited conduct.

A more pragmatically viable approach is to extend the causal analysis upstream, attributing the 'act' not to the algorithm but to the human decisions—in design, data selection, testing, or deployment—that made the harmful output foreseeable. This approach is consistent with established principles of causation in both criminal and tort law, and it avoids the conceptual difficulties associated with attributing legal agency to non-human systems. The Robodebt scheme—in which the Australian government deployed an income-averaging algorithm that generated approximately 443,000 erroneous debt notices—provides a powerful illustration: the unlawful act was not the algorithm's calculation, but the public servants' decision to deploy a system they knew, or ought to have known, was producing systematically incorrect results (Royal Commission into the Robodebt Scheme, 2023).

4.2 Mens Rea and the Problem of Algorithmic Intention

The doctrine of Mens Rea—the requirement of a guilty mind—presents an even more fundamental challenge. Criminal liability in most jurisdictions requires proof that the defendant acted with a specified mental state: intention, recklessness, or negligence. Autonomous algorithmic systems are incapable of possessing any of these states. Three doctrinal models have emerged in the literature as potential responses to this difficulty.

The 'Innocent Agent' model, developed by analogy from the established doctrine of innocent agency in principal-and-agent liability, treats the algorithm as a mere instrument of the human programmer or deployer, who retains full Mens Rea. This model is jurisprudentially conservative and consistent with existing doctrine, but it may produce unjust outcomes where the responsible human's state of mind did not encompass the specific harm that resulted.

The 'Probable Consequence' model attributes constructive intent to human operators for harmful outcomes that were objectively foreseeable at the time of deployment. Drawing on the established tort standard of the reasonable professional, this model would hold a government agency liable for algorithmic discrimination if a reasonably competent public authority would have foreseen the discriminatory impact. This approach finds support in the ECtHR's jurisprudence on positive obligations under Article 1 Protocol 1 and Article 14 ECHR.

The 'Algorithmic Intention' model—the most conceptually ambitious of the three—proposes evaluating intent at the level of the system's design architecture, treating the objective functions programmed into a model as analogous to the mental states of human actors. While this model captures an important intuition—that biased design choices represent a form of 'frozen intention'—it faces the fundamental objection that intention requires subjective mental states that code, however sophisticated, cannot possess (Duff, 2007; Floridi et al., 2018).

4.3 The Proposed Liability Architecture

The paper advances a Hybrid Liability Model that integrates elements of all three doctrinal approaches whilst avoiding their individual limitations. The model operates through a 'chain of command' structure that allocates distinct liability obligations to each actor in the algorithmic supply chain.

Developers bear primary design liability for incorporating biased training data, deploying unvalidated models, or failing to implement proportionate fairness constraints. This liability is strict in character—it does not require proof of intent—but it is limited to foreseeable categories of harm identified in a mandatory pre-deployment Fundamental Rights Impact Assessment (FRIA). Data providers are strictly liable for the quality and representativeness of training datasets where they

supply data commercially, subject to a reasonable care defence for unforeseeable sources of bias. Public officials bear personal liability in negligence where they deploy algorithmic outputs as an operational substitute for professional judgment—'rubber-stamping' decisions without substantive review—where this conduct falls below the standard expected of a reasonably competent public administrator.

ADMINISTRATIVE SAFEGUARDS AND THE EVOLVING STANDARD OF JUDICIAL REVIEW

5.1 The Right to Explanation as a Constitutional Imperative

The 'right to explanation'—the legal entitlement of individuals to receive comprehensible reasons for automated decisions affecting their legal interests—has emerged as the central procedural battleground in algorithmic governance. The GDPR's Article 22 provides the foundational legal basis in European law, conferring on data subjects the right not to be subject to solely automated decisions that produce significant legal effects, and imposing on data controllers an obligation to implement 'suitable safeguards,' including 'at least' the right to obtain human intervention, express one's point of view, and contest the decision.

The EU AI Act considerably strengthens these obligations for 'high-risk' systems, as defined in Annex III. Article 13 requires that high-risk systems be designed to ensure transparency, and Article 14 mandates meaningful human oversight. Most significantly, Article 86 establishes an explicit right to explanation for decisions made by AI systems in high-risk domains. The interpretive question—what constitutes an adequate explanation—remains contested, but the scholarly consensus, led by Selbst and Barocas (2018), holds that a legally adequate explanation must be counterfactually informative: it must identify the principal features driving the decision and indicate what the applicant could do differently to achieve a different outcome.

5.2 Judicial Review of Algorithmic Upstream Choices

Traditional judicial review doctrine focuses on the legality and rationality of administrative outputs—the final decision. This approach is structurally inadequate in the algorithmic context because it cannot reach the 'upstream choices' that determine algorithmic behaviour: the selection of training data, the weighting of variables, and the setting of decision thresholds. A more adequate standard of review would examine the algorithmic system's architecture as a continuous administrative act, treating design and deployment decisions as subject to review on the grounds of illegality, irrationality, and procedural impropriety—the three heads of review established in Council of Civil Service Unions v Minister for the Civil Service [1985] AC 374.

The four-layer analytical framework advanced in this paper operationalises this extended standard of review. Courts applying this framework would be required to examine not merely whether a specific decision was lawful, but whether the algorithmic infrastructure that produced it was designed and deployed in compliance with applicable legal obligations—including the prohibition on proxy discrimination, the requirement of explainability, and the obligation of meaningful human oversight.

PROPOSED GOVERNANCE FRAMEWORK

6.1 The Legal Risk Index (LRI)

This paper proposes the Legal Risk Index (LRI) as a standardised diagnostic instrument for assessing the regulatory sensitivity and legal vulnerability of public sector ADM deployments. The LRI

is calculated as the arithmetic mean of three component scores:

LRI = (P + L + F) / 3

Privacy Risk (P) measures the sensitivity of personal data processing against applicable data protection standards, on a scale of 1 to 10, with higher scores indicating greater processing intensity or sensitivity. Legal Risk (L) evaluates the degree to which the system's design and operation complies with applicable legal standards, including non-discrimination obligations, with higher scores indicating greater non-compliance risk. Freedom Risk (F) assesses the potential impact of the system on fundamental rights and civil liberties, including the right to a fair hearing and the right to non-discrimination, on a scale of 1 to 10.

Systems scoring above 7.0 on the LRI should be treated as 'high-risk' for the purposes of the FRIA obligation and should trigger enhanced transparency requirements, mandatory human oversight mechanisms, and strict liability in the event of harm. The LRI is not intended as a definitive determination of legal compliance but as a structured analytical tool that public authorities and their legal advisers can use to prioritise oversight resources and identify areas requiring pre-deployment remediation.

6.2 The Fundamental Rights Impact Assessment (FRIA)

Modelled on the environmental impact assessment tradition and drawing on the EU AI Act's Article 9 requirements, the FRIA proposed in this paper constitutes a mandatory pre-deployment review process for all government ADM systems that score above 6.0 on the LRI. The FRIA comprises five sequential stages.

The first stage is system characterisation, which requires a comprehensive technical description of the system's architecture, including its training data sources, algorithmic methodology, and decision outputs. The second stage is legal compliance audit, which involves systematic assessment of the system against applicable legal obligations, including GDPR Article 22, EU AI Act Annex III, and relevant domestic data protection legislation. The third stage is bias and discrimination testing, which requires empirical testing of the system's outputs across relevant demographic groups to identify disparate impact, using validated statistical methods including demographic parity analysis and equalised odds testing. The fourth stage is explainability assessment, which evaluates whether the system is capable of generating counterfactually informative explanations of sufficient quality to satisfy the right to explanation under applicable law. The fifth stage is human oversight design, which specifies the mechanisms through which human reviewers will exercise substantive—rather than nominal—oversight of algorithmic outputs, including the training, authority, and accountability of reviewers.

SECTORAL CASE STUDIES: ALGORITHMIC BIAS IN CRITICAL DOMAINS

7.1 Criminal Justice and Predictive Policing

The use of algorithmic risk assessment tools in criminal justice represents perhaps the most jurisprudentially consequential deployment of ADM technology. The COMPAS system, deployed across numerous US states for pre-trial detention and sentencing decisions, became the subject of the landmark Wisconsin Supreme Court decision State v. Loomis, 881 N.W.2d 749 (Wis. 2016), in which the court upheld the use of COMPAS scores in sentencing whilst requiring that they not constitute the determinative factor in the judge's decision. The court's reasoning implicitly acknowledged the constitutional concern that a defendant cannot meaningfully challenge a sentence informed by a proprietary algorithm whose methodology is a protected trade secret.

Subsequent empirical research by Dressel and Farid (2018) demonstrated that the COMPAS algorithm's predictive accuracy was no better than that of untrained lay volunteers completing a brief online survey—a finding that fundamentally undermines the epistemic justification for algorithmic risk assessment tools that attract heightened procedural protections precisely because of their claimed predictive superiority.

In the United Kingdom, the Law Society's Technology and Law Policy Commission (2019) called for a National Register of Algorithmic Systems deployed in the justice system, arguing that the absence of centralised oversight creates an unacceptable accountability deficit. This proposal reflects the view, advanced in this paper, that algorithmic governance requires systemic rather than merely case-by-case accountability mechanisms.

7.2 Administrative Welfare and the Robodebt Paradigm

The Australian Robodebt scheme provides perhaps the most extensively documented case study of algorithmic governance failure at scale. Between 2016 and 2019, the Australian Department of Human Services deployed an automated income-averaging algorithm to identify welfare recipients whose reported earnings differed from tax office records, automatically generating debt notices without individual case assessment. The Royal Commission into the Robodebt Scheme (2023) found that the scheme was unlawful from inception because the methodology of income-averaging as a basis for raising a debt had no legal foundation in the applicable social security legislation.

The Robodebt case illustrates the central argument of this paper with particular clarity. The 'act' that produced harm was not the algorithm's calculation—which performed precisely as designed—but the decision by senior public servants to deploy a system they knew, or ought to have known, produced legally indefensible outcomes. The Robodebt Commission's recommendation for criminal referrals in relation to senior officials who misled parliamentary committees reinforces the proposition that algorithmic harm can and should generate personal liability for responsible human actors.

7.3 The Dutch SyRI Case

In a landmark ruling of 5 February 2020, the Hague District Court found that the Dutch government's System Risk Indication (SyRI) algorithm—which combined data from fourteen government sources to generate risk scores identifying individuals likely to commit welfare fraud—violated Article 8 ECHR (Rechtbank Den Haag, 2020, ECLI:NL:RBDHA:2020:1878). The court found that the system's opacity prevented effective judicial review and that its disproportionate deployment in low-income urban districts disclosed a pattern consistent with indirect discrimination on grounds of socioeconomic status.

The SyRI judgment is significant for three reasons. It establishes that the opacity of an algorithmic system can independently constitute a violation of Convention rights, regardless of whether any specific decision produced an unjust outcome. It demonstrates that disparate impact analysis—rather than proof of discriminatory intent—is a legally sufficient basis for challenging algorithmic governance systems. And it confirms that judicial review can appropriately extend to the upstream design and deployment choices that determine a system's behaviour, rather than being confined to the legality of individual outputs.

THE SAUDI ARABIAN AND GCC REGULATORY LANDSCAPE

8.1 Vision 2030 and the SDAIA Framework

Saudi Arabia occupies a distinctive position in the global landscape of algorithmic governance: a state committed to rapid AI-led digital transformation and simultaneously seeking to develop robust

ethical and legal guardrails for that transformation. The Saudi Data and Artificial Intelligence Authority (SDAIA), established by Royal Decree in 2019, has played a central institutional role in this endeavour, publishing the National Strategy for Data and AI, the Saudi AI Ethics Principles (2022, revised 2025), and administering the PDPL framework.

The 2025 revisions to the Kingdom's Ethical AI Principles are particularly significant for the purposes of this paper. The revised Principles impose explicit obligations of fairness, accountability, and transparency on ADM systems deployed in public registries, healthcare, and the justice sector—domains that correspond directly to the 'high-risk' categories identified in the EU AI Act's Annex III. This regulatory convergence reflects a conscious policy decision by SDAIA to align Saudi AI governance with international best practice while maintaining sovereignty over implementation.

8.2 The PDPL as a Constitutional Buffer

The Personal Data Protection Law (PDPL), promulgated by Royal Decree M/19 on 9 September 2021 and substantively amended in 2023, establishes the primary legal framework for data protection in the Kingdom. Article 25 PDPL—analogous in structure to GDPR Article 22—provides that automated processing decisions that significantly affect data subjects must incorporate mechanisms for human review and must be capable of explanation upon request.

However, the PDPL's provisions in this area exhibit several important limitations relative to European instruments. The right to human review under Article 25 is triggered only by decisions that 'significantly affect' data subjects, a threshold that remains undefined in both the law and implementing regulations. The PDPL does not impose an obligation equivalent to the EU AI Act's requirement for a Fundamental Rights Impact Assessment before deployment of high-risk systems. And the enforcement mechanism—supervision by the National Data Management Office—lacks the independence and investigatory resources necessary for effective algorithmic accountability.

This paper recommends that Saudi Arabia address these limitations through targeted amendments to the PDPL and its implementing regulations, drawing on the EU AI Act's risk-tiered approach. Specifically, the implementing regulations should define the threshold for 'significant effect' by reference to an exhaustive list of high-risk domains; require FRIA for all government ADM systems scoring above 6.0 on the LRI; and establish an independent algorithmic audit function within SDAIA with the power to issue binding compliance orders.

8.3 Regional Harmonisation and the 'Third Way' Model

Across the GCC, regulatory approaches to algorithmic governance vary considerably in sophistication. The UAE's Model AI Governance Framework (2023) represents the most detailed regional instrument, establishing sector-specific risk assessment requirements and an AI certification system for government deployments. Qatar's National AI Strategy (2022) emphasises sovereign data governance but provides limited procedural protections for individuals affected by automated decisions. Bahrain, Kuwait, and Oman have not yet developed specialised AI governance instruments.

This regulatory fragmentation creates a significant impediment to cross-border algorithmic accountability in a region characterised by deep economic integration. A company deploying an ADM system across multiple GCC jurisdictions faces inconsistent legal obligations and the risk of simultaneously satisfying one jurisdiction's requirements whilst violating another's. The paper recommends the negotiation of a GCC Framework Convention on Algorithmic Governance, modelled on the Council of Europe's Convention 108+, that would establish common minimum standards for transparency, impact assessment, and individual redress whilst preserving member states' flexibility in

implementation.

DISCUSSION

The analysis presented in this paper yields several findings of broader significance for the jurisprudence of algorithmic governance. First, the doctrinal inadequacy of existing criminal law concepts—particularly Mens Rea—in the algorithmic context is not a peripheral problem but a structural one. The assumption of human agency that underpins criminal liability cannot be supplemented by legal fiction or analogical extension; it requires a deliberate legislative response that acknowledges the distributed, multi-actor character of algorithmic harm.

Second, the Hybrid Liability Model advanced here demonstrates that the responsibility gap is not inherently uncloseable. By allocating distinct liability obligations to developers, data providers, and deploying officials—calibrated to their respective epistemic positions and decision-making authority within the algorithmic supply chain—a legally coherent accountability framework can be constructed without requiring attribution of legal personality to algorithms themselves. This approach is more jurisprudentially conservative than proposals for 'electronic personhood,' and is more likely to command judicial acceptance in jurisdictions committed to established doctrinal categories.

Third, the FRIA and LRI instruments proposed in this paper advance the practical implementation of algorithmic accountability by providing structured analytical tools that public authorities can apply before deployment. Proactive governance instruments of this kind are superior to post-hoc remediation in the algorithmic context because the harm from biased systems typically accrues to large populations through numerous individual decisions before any single case achieves legal salience.

Fourth, the comparative analysis of the EU, Australian, and Saudi Arabian regulatory experiences demonstrates that effective algorithmic governance does not require a single regulatory model. The EU's precautionary rights-based approach, Saudi Arabia's values-driven developmental model, and the post-hoc accountability mechanisms employed in Australia each offer distinct advantages and limitations. What they share—and what the comparative literature confirms is essential—is an explicit legislative commitment to the principle that automated state power must remain subject to meaningful human oversight and judicial review.

CONCLUSION AND POLICY RECOMMENDATIONS

10.1 Synthesis of Findings

This paper has examined the 'responsibility gap' generated by the deployment of autonomous algorithmic systems in public administration, with particular attention to the doctrinal challenges these systems pose for criminal liability and procedural due process. The analysis has demonstrated that existing legal frameworks—designed for human actors operating within transparent decision-making processes—are structurally inadequate to the governance of algorithmic systems characterised by opacity, distributed agency, and emergent behaviour.

The paper has advanced three original contributions to address this inadequacy. First, a Hybrid Liability Model that distributes responsibility across the algorithmic supply chain whilst preserving the doctrinal integrity of criminal law concepts. Second, a Legal Risk Index that provides a standardised instrument for assessing regulatory vulnerability and prioritising oversight. Third, a FRIA framework that operationalises the principle of preventive algorithmic accountability as a design-level obligation.

10.2 Policy Recommendations

For legislators and regulators, the paper advances the following recommendations. First, enact mandatory algorithmic impact assessment requirements for all government ADM systems operating in high-risk domains, with independent audit and enforcement mechanisms. Second, adopt a statutory right to a meaningful algorithmic explanation, defined by reference to a counterfactual informativeness standard, and cognisable before administrative tribunals. Third, establish a hybrid liability framework that allocates design liability to developers, data liability to providers, and operational liability to deploying officials, with a rebuttable presumption of strict liability for high-risk systems. Fourth, GCC states should negotiate a regional framework convention establishing common minimum standards for algorithmic governance and creating a joint supervisory body with powers of investigation and enforcement.

10.3 Limitations and Future Research Directions

Several limitations of the present analysis should be acknowledged. The normative juridical method, whilst appropriate for doctrinal inquiry, does not generate empirical data on the real-world performance of the liability frameworks examined. Future research should combine doctrinal analysis with empirical study of how algorithmic accountability frameworks operate in practice—including how the proposed FRIA is implemented, how courts assess LRI scores, and how the Hybrid Liability Model performs in contested litigation.

Additionally, this paper's comparative focus on the EU, Australia, and Saudi Arabia necessarily omits important regulatory developments in other jurisdictions, including China's Provisions on the Management of Algorithmic Recommendations (2022) and the United States' Blueprint for an AI Bill of Rights (2022). A systematic global comparative study of algorithmic governance frameworks would constitute a valuable contribution to the field and is identified as a priority for future research.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31

Council of Civil Service Unions v Minister for the Civil Service [1985] AC 374 (House of Lords, United Kingdom) .

Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62. https://doi.org/10.1145/2844110

Diab, M. F. S. (2024). Personal data protection and AI governance in the GCC: Comparative regulatory analysis. Arab Law Quarterly, 38(1), 1–34. https://doi.org/10.1163/15730255-bja10096

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., … & Wood, A. (2017). Accountability of AI under the law: The role of explanation. Berkman Klein Center for Internet & Society Working Paper. https://ssrn.com/abstract=3064761

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

Duff, R. A. (2007). Answering for crime: Responsibility and liability in the criminal law. Hart Publishing.

Edwards, L., & Veale, M. (2022). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. Duke Law & Technology Review, 16(1), 18–84.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St Martin's Press .

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88 .

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689 .

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2018). How to design AI for social good: Seven essential factors. Science and Engineering Ethics, 26(3), 1771–1796. https://doi.org/10.1007/s11948-018-0063-0

Law Society of England and Wales. (2019). Technology and the law: Policy commission final report. The Law Society.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology, 6(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Rechtbank Den Haag. (2020, February 5). ECLI:NL:RBDHA:2020:1878 (SyRI judgment). Rechtbank Den Haag.

Royal Commission into the Robodebt Scheme. (2023). Report of the Royal Commission into the Robodebt Scheme. Commonwealth of Australia. https://robodebt.royalcommission.gov.au/publications/report

Saudi Data and Artificial Intelligence Authority (SDAIA). (2022). National strategy for data and AI. Kingdom of Saudi Arabia. https://sdaia.gov.sa

Saudi Data and Artificial Intelligence Authority (SDAIA). (2025). Ethical AI principles (revised). Kingdom of Saudi Arabia.

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. Fordham Law Review, 87(3), 1085–1139.

Solum, L. B. (1992). Legal personhood for artificial intelligences. North Carolina Law Review, 70(4), 1231–1287.

State v. Loomis, 881 N.W.2d 749 (Wis. 2016). Wisconsin Supreme Court, United States.

Turner, J. (2019). Robot rules: Regulating artificial intelligence. Palgrave Macmillan.

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). 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. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887.

Zarsky, T. Z. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology, & Human Values, 41(1), 118–132. https://doi.org/10.1177/0162243915605575