SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Algorithmic State Power And Constitutional Democracy: Reimagining The Rule Of Law In The Age Of Artificial Intelligence And Machine Learning

Avatar Chaubey[1*], Akanksha Verma[2], Deepika Dhiman[3], Neha Khanna[4], Ridhipa Jakhar[5], Deepika Kanwar[6], Awanit Kumar[7]

[1,2,3,4,5,6]Assistant Professor, Amity Law School, Amity University Punjab, Mohali, India
[7]Associate Professor, Department of CSE, Career Point University, Kota, Rajasthan, India
Email: 277chaubey1995@gmail.com[1*], akankshaverma.law@gmail.com[2], deepika.dhiman1991@gmail.com[3], neha.khanna1312@gmail.com[4], ridhipa.j@gmail.com[5], advdeepikakanwar1@gmail.com[6], itsawanitkumar@gmail.com[7]

Abstract- AI and machine learning become more and more integral to the decision-making process within the public sector, important constitutional issues arise about transparency, explainability, accountability, fairness, privacy and procedural due process. The aim of this study is to fill the gap in the literature on systematically and interpretable ways to judge if algorithmic state power is in accordance with constitutional democracy and the rule of law. The goal is to create a quantitative, interpretable model for detecting and assessing constitutional risks in government algorithmic systems. The semantic representation of legal entities with Legal-BERT, the extraction of constitutional information and the implementation of a neuro-symbolic rule engine enable the analysis of records within the UK Algorithmic Transparency Recording Standard (ATRS) repository in the proposed Constitutional Neuro-Symbolic Risk Assessment (CNSRA) framework. The CARI is a metric that measures governance risks identified. The analysis reveals the following concerns with transparency (73.33%), explainability (70.00%), accountability (68.33%), due process (65.00%), fairness (60.00%) and privacy (56.67%). The overall CARI component score is 77.27% and the highest domain-level risk scores are found in law-enforcement applications (82.40%). The novelty of the approach is to combine legal language intelligence with the explicit constitutional reasoning and quantitative risk measurement. The framework outlines a readable and explainable audit framework of algorithmic state power, and concludes that constitutionally protecting rights is a requirement for a more transparent, accountable, rights-respecting and democratic approach to governance.

Keywords—Algorithmic State Power; Constitutional Democracy; Artificial Intelligence; Constitutional Neuro-Symbolic Risk Assessment (CNSRA); Algorithmic Transparency; Constitutional Algorithmic Risk Index (CARI)

I. Introduction

A. Algorithmic State Power and AI-Driven Public Decision-Making

In an increasingly artificial intelligence and machine-learning-driven world, public administrations are relying more and more on these systems to make consequential judgments, such as screening for welfare eligibility, predictive policing, or immigration risk scoring. The shift of decision-making power from human officials to statistical models is a new type of algorithmic state power, a power that is coercive and distributive in nature and has traditionally been in the hands of constitutionally accountable officials. Police the draft and predict the impact of the law: Legislatures are starting to use algorithmic support in law-making as well. Predict the impact of the law, and police the draft: Legislatures are beginning to experiment with algorithmic support in the process of drafting the law. As algorithms become the mediators between citizens and the state, the quality, consistency and interpretability of the laws and technical terms governing them gain in importance in order to

uphold the rule of law [2]. Many machine learning systems, especially deep neural networks, are opaque, which limits the ability of impacted people to explain to themselves why a decision was made, thus violating the obligations in the context of reason-giving, which have long been the basis of the legitimacy of administration. Algorithmic state power is not the same as previous forms of bureaucratic discretion, as it is executed on a large scale, with little human oversight or attention, and in some instances, using poorly documented or proprietary algorithms where a conventional oversight system can't be applied. This change is not just a change in technique but a change in structure, a change in the distribution, exercise and control of power in the modern administrative state. The long-term trend of changing political and administrative structures can already be seen, as automation takes over what humans traditionally do, and where sovereignty and discretion will lie in the future [3]. In democratic institutions, automated decision-making can end up replacing technical optimisation with a decision that is based on the constitution, which has significant consequences for citizens' rights and fair treatment expectations [4]. As AI is increasingly being used to make decisions on access to public goods and services, like fraud detection, resource allocation, and risk classification, it becomes a potent tool for the state to control people.AI is being used more and more in making decisions regarding access to public goods and services, including fraud detection, resource allocation and risk classification, and these are becoming powerful tools for the state to control people. The constitutionally assumed transparency and reviewability of state action is on a knife's edge in these systems, which are often bought from private vendors and protected by trade-secret laws. A discussion of the power of algorithms thus needs to be interdisciplinary, linking computer science, public law and democratic theory to examining technical decisions in the context of a broader normative architecture of constitutional governance

B. Constitutional Democracy and Rule-of-Law Challenges

Transparency, accountability, proportionality, non-discrimination and the right to a fair hearing are key principles of constitutional democracies, which are undermined by decisions being made behind a closed door with the help of opaque computational systems. But many algorithmic systems fail to meet all of these expectations: the rule of law requires that state action be predicated on a clear legal basis, meaning it is described in understandable terms; and has to be subject to challenge by means that are easily accessible. AI tools are also influencing how representative institutions are formed; should the priorities and participation that they drive replace that of citizens in determining the nature of democratic representation? Digital intermediaries and automated content and decision systems more than ever raise fundamental rights issues, and we need to pay more attention to fundamental rights regimes and make them a priority rather than a compliance issue. The additional challenges of explainability deficits are exacerbated by the fact that a citizen who is denied a benefit or identified as a high-risk by a machine-learning model is unlikely to be able to obtain a reasoned account to make a meaningful legal challenge and thus undermines the notion of procedural due process at the heart of constitutional adjudication [5].

Accountability systems for human decision makers such as administrative appeals, judicial review, and ombudsman overviews were not developed to question statistical models, gradient-based classifiers or large training corpora of decisions that influence the decisions made by such systems. Fairness issues relate to the fact that the training data that is used to train machines may contain historical patterns of discrimination, which may be learned by automated systems and can be scaled up in negative ways that are likely to affect already marginalised communities [6]. Similarly, privacy is involved, because many of the algorithmic state systems rely on the aggregation of large amounts of personal data, far greater than what was thought of when privacy laws were enacted to incorporate data-protection principles. While the design of democratic institutions in an algorithmic world is about the changing nature of citizen engagement in governance, it is also about how constitutional checks can be incorporated into the systems, as well as how they can be applied upon deployment. Re-imagining democratic institutions for an algorithmic world is about the changing nature of citizen engagement in governance, but it is also about the design of constitutional checks themselves which can be embedded in the systems and not just applied after deployment. Therefore, the rule-of-law challenge is not just to control AI after it has been already developed but to integrate the rule of law, its values, principles and ideals (such as legality, proportionality and contestability) into the design, procurement, and operation of algorithmic systems. The challenge is met by using interpretable and quantifiable methods that are able to systematically identify the gaps between algorithmic state power and constitutional expectations this study aims at bridging this gap by providing a structured and evidence-based risk-assessment framework.

C. Research Problem, Objectives, and Motivation

Research Problem:

Although a fair amount of scholarly research has appeared on AI governance, there is no systematic and interpretable way to measure the constitutionality of the algorithmic state system and whether it meets the principles of constitutional democracy and rule of law. Existing disclosures of transparency are predominantly descriptive, with few if any structured constitutional reasons or measurable risk signals, leaving regulators, courts and citizens with no reliable evidence base to consistently, comparably and meaningfully critique algorithmic power in the state across various application domains in the public sector.

Objectives:
This study is going to develop and test a quantitative and explainable model to derive constitutional risk signals from transparency records of government algorithms. The system aims to perform legal and technical disclosures with the help of Legal-BERT embeddings, identify constitutional entities and constitutional safeguards through the neuro-symbolic reasoning, and calculate an aggregate Constitutional Algorithmic Risk Index which would allow for comparative auditing across different domains of public-authorities.

Motivation:
The impetus is fuelled by the growing disconnect between the quick adoption of algorithms by the government and the comparatively slow development of constitutional oversight mechanisms, which are able to examine such systems. In the era of algorithmic decision making affecting liberty, welfare and due-process rights, and an audit methodology that is interpretable and reproducible is just as necessary as a way to preserve the democratic legitimacy of the algorithmically-driven state action, protect fundamental rights, and restore public confidence in algorithmic state action.

D. Major Contributions of the Study

- ✓ Innovatively proposes a Constitutional Neuro-Symbolic Risk Assessment (CNSRA) framework integrating Legal-BERT semantic and explicit rule-based constitutional reasoning.
- ✓ Proposes a repeatable, algorithmic state power auditable metric – the Constitutional Algorithmic Risk Index (CARI).
- ✓ Empirically assesses 60 records of ATRS for the United Kingdom and finds transparency, explainability and accountability to be the prevailing dimensions of the constitution as a risk.
- ✓ Provides a comparative risk analysis on a domain level, between the risk of law enforcement, welfare, immigration and algorithmic applications in administration.

II. Related Work and Research Gap

The use of artificial intelligence by governments in administrative decisions has grown exponentially in the areas of tax, welfare, immigration and law enforcement, leading to an investigation of the possible ways that legal systems can incorporate AI-based administrative decisions without sacrificing important doctrinal protections [7], [8]. Reacting to AI through the legal lens has been a lagging aspect of AI implementation, with courts and legislatures typically responding to AI harms only after there is a strong public debate about them. The argument for conceptual plans of adaptive regulation suggests that rules need to be technology agnostic, meaning that they need to adapt to the evolution of AI technology, instead of being technology specific and therefore dangerous to be mismatched with the changing models and their context of use [9]. The conflict between boosting the efficiency of algorithms and protecting rights within the boundaries of the law has been described as a ‘fundamental rights clash': what is required from public governance is to balance computational efficiency with constitutionally protected rights like the right to equality, the right to due process, etc. [10].
These strands of literature all agree that algorithmic decision-making in government is not simply a 'technical fix' that enhances the current way government decisions are made but a structural 'reconfiguration' of how government power is used, recorded and evaluated. The timelines for decisions are shortened, the individual case reasoning is hidden, and the accountability for the decision is spread across numerous vendors, data sources and versions of models, all of which impact traditional principles of ministerial responsibility. Enabling agencies to automate their processes consistently highlights that there is a need for the process to be both procedural and substantive faithful to the law, which is incorporated in the design and training inputs of the algorithm. In addition, the literature reports that public bodies often do not have the knowledge and skills to challenge the models that are supplied by the vendors, and, even when models are explicitly set out in the transparency laws, this is not achieved.

The literature also notes that there is often an unequal knowledge level between the vendors and the public bodies, which hinders effective oversight even where transparency obligations are formally set out in the transparency laws. The imbalance is especially pronounced in the context of criminal justice and border control, where the stakes of individual freedom are greatest, and the reasons why such opacity is easier to come by and more profitable commercial and security-related – are most plentiful. A pattern that emerges from a range of comparative studies of AI governance in democratic contexts is that there is significant variation regarding the balance between the promotion of innovation and rights protection, but that algorithms' contribution to public governance ultimately calls into question the need for legal structures that enable tracking of the outputs of algorithms to the identifiable, contestable legal reasoning that produced them. The lack of this traceability puts algorithmic efficiency at risk of being legally and democratically unaccountable; therefore, the need for the development of structured assessment tools that can systematically assess governments' algorithmic systems against set legal and constitutional standards, instead of ad hoc or purely technical audits.

B. Constitutional Rights, Transparency, Fairness, and Accountability

Algorithmic bias, opacity and lack of accountability are becoming a growing legal and constitutional issue, not only on the conceptual level, but also through the formulation of new laws designed to govern real functioning algorithms. Algorithmic bias is attributed to various factors including the challenges of training an algorithm from non-representative data, the use of proxy variables that are related to protected attributes and feedback loops that may reinforce already-existing discriminatory patterns, which all depend on constitutional equality guarantees [11]. When leveraged in a context where the digital transformation is fast, AI optimisation models can raise governance risks that go beyond issues of fairness, to systemic issues of market stability and institutional accountability in financial and risk-prediction use-cases [12]. AI in Criminal Procedure Scholarship focuses on certain areas of due process and evidence related to the use of AI in investigative and prosecutive processes, such as whether an individual should be released, whether a defendant should be detained, and whether a defendant's sentence is to be extended.

Transparency is frequently expressed in administrative law, but can be ineffective to assure explanations because the provided disclosure of the system may not lead to explanations of the decisions made. The individual accountability structures struggle with the decentralized accountability in the data scientists, procurement officials, and third parties involved, where they might all play a minor role, but in the end may have contributed to the cause of the harm from an algorithm. Fairness analysis has been done more than statistical equality and substantive equality of the Constitution, since formally neutral algorithms can have different effects on race, gender and/or socio-economic status. As the data used for training and running government algorithms is often not subject to purpose-limitation restrictions found in data-protection law, these accountability shortcomings and privacy issues combine to increase the potential risk faced by those impacted. There are a number of contributions that suggest the need to re-interpret or supplement the current constitutional interpretations or doctrines, which have been developed to govern human decision makers in “bounded administrative” contexts where they are allowed discretion. As a whole this body of work demonstrates that transparency, fairness, and accountability are not stand-alone constructs, but rather work with each other dynamically: flaws in one dimension (e.g., limited explainability) have a direct effect on the viability of the protection in another dimension (e.g., the right to contest an adverse decision).

C. Explainable and Responsible AI for Government Systems

While there have been many tools and bodies of work that have paved the way for explainable and responsible AI, not many of these tools are well suited for government accountability practice or the legal reasoning process. While technical approaches to explainability of AI systems are useful, they are not a sufficient basis for legal criteria of reasoned decision making as court and administrative reviewers will need explanations which focus on relevant legal considerations not just upon what is important, but upon what those things are important for. This conflict is perfectly captured in judicial use of AI, where there are proposals for ethical guidelines on the use of AI that call for “meaningful interpretive discretion” for judges and other decision makers even when making decisions with the aid of an AI tool; and the idea that judgements made by AI tools must inform the reasoning of the judges, but cannot replace it [14].

The governance models of the Internet-of-Things and Industry 4.0, which we have elaborated first for these two application areas, can also be transferred to the public sector when it comes to AI governance: Socio-legal governance layers are put in place on top of technical and organizational governance layers to keep control of the systems while allowing for innovation to flourish [16]. There are multiple principles that are typically part of responsible AI frameworks for government systems such as human oversight, contestable and proportional

automation and translating these principles into measurable and auditable requirements is still a methodological challenge. There are also other types of algorithmic personalisation and ranking systems, like recommendation and service-discovery systems, which could not always be formulated in a constitutional way, but which can still affect the access to public information and services for which accountability should be looked at more closely [18].

D. Existing Limitations and Identified Research Gap

A common limitation found in these three streams is the lack of comprehensive literature on the three aspects of algorithmic governance, constitutional rights and explainable AI, particularly one that is holistic, uniform and can be applied repeatedly to evaluate these three in a single, quantifiable way that can measure constitutional risk. The commentaries on the law and policy are mostly qualitative, doctrinal and (normative) analytical, but do not necessarily quantify to ensure that the indicators can be applied in a consistent way to a large number of government algorithmic systems.

The proposed CNSRA framework and the related Constitutional Algorithmic Risk explicitly address this gap, by combining the language of science (Legal-BERT) with a symbolic constitutional rule engine, so as to connect the qualitative-doctrinal approach with the quantitative-technical approach that were identified in the reviewed literature.

Table 1. Summary of Related Work on Algorithmic Governance, Constitutional Rights, and Explainable AI

Study	Focus Area	AI/Algorithmic Domain	Key Constitutional Dimension	Methodology	Key Finding
[8]	Legal reactions to AI	General public/private AI use	Rule of law adaptation	Doctrinal analysis	Legal responses to AI remain largely reactive and lag deployment
[9]	Adaptive AI regulation	Cross-sectoral AI systems	Legality & foreseeability	Conceptual framework	Static rules become obsolete; adaptive regulation needed
[10]	Algorithm power vs. rights	Public governance AI	Rights conflict & governance	Legal-governance analysis	Algorithmic efficiency conflicts with legal-boundary protections
[11]	Algorithmic bias	Automated decision systems	Equality/non-discrimination	Conceptual & regulatory review	Bias originates from data, proxies, and feedback loops
[12]	AI risk prediction	Financial risk management	Institutional accountability	Optimized graph-network model	AI risk models raise systemic accountability concerns
[13]	AI in criminal procedure	Criminal justice/law enforcement	Due process	Legal-practice case study	AI integration raises due-process and evidentiary challenges
[14]	AI in judicial administration	Judicial decision-making	Fair hearing & judicial reasoning	Ethical-framework proposal	Judges must retain interpretive discretion over AI outputs

[15]	AI regulation overview	Cross-sectoral AI	Explainability & compliance	Regulatory survey	Technical explainability alone insufficient for legal standards
[16]	Socio-legal IoT/Industry 4.0 governance	IoT/Industry 4.0 systems	Layered accountability	Middle-out/inside-out model	Layered governance preserves accountability without stifling innovation
[17]	Human-rights-based AI governance	Public-sector AI (Europe)	Rights-based explainability	Normative framework	Explainability should be anchored in enforceable rights obligations

In this table 1, we have compiled 10 representative contributions that are relevant to algorithmic state power in legal, technical and governance aspects. The comparison highlights that there is a perceived disconnect between the anticipation, explainability and rights-based focus, which is an identified gap. Most studies focus on doctrinal issues or technical benchmarking, while few offer any type of integrated and measurable tools for the assessment of transparency records in the framework of constitutional auditing, at large scale.

III. ATRS Dataset and Data Preparation

A. UK Algorithmic Transparency Recording Standard (ATRS) Repository

The UK Algorithmic Transparency Recording Standard (ATRS) is a public repository of structured information on algorithmic tools that public authorities and departments use in deciding on actions that impact on citizens, which is maintained by the government. A record for each tool outlines the purpose of the tool, the public authority responsible, the decision-making context, data inputs, oversight mechanisms and risk mitigation measures. It includes a diverse and yet well-defined collection of documents covering various areas of welfare administration, law enforcement, immigration control, fraud detection, and public recruitment, appropriate to systematic constitutional risk analysis of different algorithmic applications of the state across different levels of algorithmic maturity.

B. ATRS Record Selection and Eligibility Criteria

The three criteria used to select records for inclusion were: that the records had to be complete with the structured disclosure fields, that the records had to have enough free text description to be used with Legal-BERT for semantic analysis, and that the records had to be relevant to decisions that had direct legal, financial, and liberty consequences for the individuals involved. Documents whose sole use is to automate internal administration, like document routing, but not affecting citizens, were excluded. After this filtering process, 60 eligible records remained, covering six application domains, with sufficient comparability between the six different application domains for domain-level analysis.

C. Structured and Unstructured Attribute Extraction

Downstream constitutional analysis was facilitated by breaking down each ATRS record into structured and unstructured (attributes) categories. Structured attributes are attributes that are also contained in the standardized ATRS form fields and are extracted into a relational schema. Examples of structured attributes are public authority name, application domain, deployment status, and stated oversight mechanism. Unstructured attributes include free text narrative elements that explain the purpose of the tool, data sources used, human review processes and risk mitigation strategies. The input in the semantic representation stage of Legal-BERT was tokenized and segmented into the clauses in the constitution that are relevant to the narrative, such as statements about human oversight, appeal rights, or safeguards for data protection. This dual extraction approach both maintains the quantitative comparability of the records and includes the legal nuance retained in the text, which is necessary for constitutional risk annotation, in a single unified analytical pipeline, combining categorical filtering with deep contextual language understanding.

D. Legal Text Preprocessing and Constitutional Annotation

For legal text preprocessing, the legal-domain tokenization, stopword retention (including negation and modals), and lemmatization were carried out, taking into account the requirement of maintaining the constitutional terminology, e.g., "due process" or "proportionate. The six dimensions of the constitution were used as the basis for the controlled vocabulary mapping from which constitutional annotation was created. Represent a record as a series of clauses; annotate each clause to a constitutional-risk category by employing similarity thresholds as follows.

$$C(x) = \text{argmax}_{\{k \in K\} \text{sim}(E(x), R_k)}, \text{ where } \text{sim}(a, b) = \frac{a \cdot b}{\|a\| \|b\|} \quad (1)$$

This equation defines the assignment of each clause x to the category k of the constitutional clause to which it has been extracted, by computing the cosine similarity between the reference embedding R_k and x. This is a semantic nearest-category classification approach to the constitutional annotation problem.

$$x' = \text{Lemmatize}\left(\text{Tokenize}\left(\text{Lower}(x)\right)\right) \backslash S, \; S = \text{StopWords} \backslash \text{NegModal} \quad (2)$$

It is a definition of the preprocessing pipeline for which lowercasing, tokenization and lemmatization are applied, but common stop-words are not. Also, certain words that are important for the legal meaning, such as negation and modals, are explicitly kept.

$$w_x = TF(x) \cdot \log\left(\frac{N}{DF(x)}\right) \quad (3)$$

This term-weighting equation calculates "term importance" at the level of each term, based on IDF, so that a rare term in the constitution will have a higher weight when it is used for annotation.

IV. Proposed Constitutional Neuro-Symbolic Risk Assessment Methodology

The proposed Conceptual Neuro-Symbolic Risk Assessment (CNSRA) framework is a computational neural symbolic five-stage pipeline that transforms the unstructured ATRS transparency records to a quantitative and interpretable risk profile. The first step is to legally preprocess and constitutionally annotate eligible records with a set of clause-level tags that are associated with six constitutional dimensions. Second, a Legal-BERT encoder is used to learn context-specific legal representations of each of the annotated clauses, which are legal-specific but away from generic language representations. Third, a constitutional information-extraction module extracts entities, risk indicators and safeguard mentions from these embeddings, transforming continuous embeddings into discrete, symbolically interpretable facts. Fourth, a neuro-symbolic rule engine is used to encode expert rules, specified in the constitution, to the facts, creating dimension-specific risk scores (on a severity scale from 1 to 5) with full traceability back to therules that led to the risk scores. Fifth, these dimension scores are combined into a single easily interpretable percentage-based score in the Constitutional Algorithmic Risk Index, allowing record and domain level comparisons. It is designed to explicitly link each classification of risk to legal criteria and source text and not just an opaque statistical inference a pattern-recognition strength combined with transparency and auditability of the explicit symbolic rules.

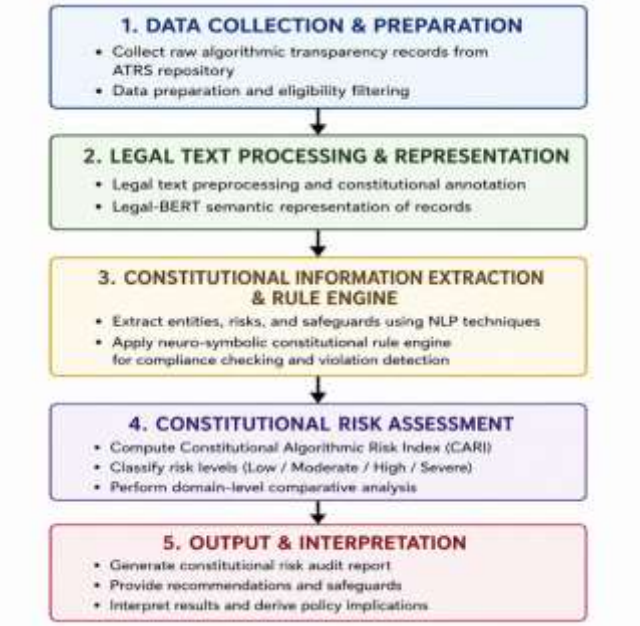

Figure 1. End-to-end workflow of the proposed Constitutional Neuro-Symbolic Risk Assessment (CNSRA) framework

Figure 1 shows the end-to-end CNSRA pipeline which first involves filtering eligibility, legal preprocessing, Legal-BERT representation, constitutional information extraction, next, evaluation of the neuro-symbolic rules, and finally, the computation of CARI and comparative reporting of rules. All outputs at each stage are structured and auditable, and feed into the next stage, leading to end-to-end interpretability of the whole constitutional risk-assessment process.

A. Legal-BERT-Based Representation and Constitutional Information Extraction

The legal embeddings produced by Legal-BERT are more accurate at capturing the domain-specific legal semantics in the context of legal documents in the ATRS transparency disclosures, compared to general purpose language models, which are trained on more general corpora. The clauses are then encoded in a fixed dimensional vector representation, and this representation is passed on to the encoder to obtain a vector representation for each preprocessed clause that is used for constitutional information extraction.

$$h_x = \text{LegalBERT}(x'), \qquad h_x \in \mathbb{R}^d$$

This equation is used to obtain the contextual embedding operation of each preprocessed clause x' into a dense vector h_x of dimension d that captures the meaning of the clause in a way that is appropriate for the domain.

$$z_x = W \cdot h_x + b, \qquad z_x \in \mathbb{R}^m$$

This linear projection maps the Legal-BERT embedding to a 6-dimensional space, where each dimension represents a constitutional risk dimension, to form a constitutional-feature space.

$$p_x = \text{softmax}(z_x)$$

The softmax function transforms the projected feature vector to a probability distribution for different constitutional risk categories which makes it possible to classify each clause probabilistically as an entity and the risk type.

$$R_{\text{score}(x)} = \Sigma_k p_{x[k]} \cdot s_k$$

The following weighted-sum equation calculates the constitutional risk score for a clause, based on the probabilities of the categories, and the severity weights s_k which are specified for each constitution dimension.

C. Constitutional Entity, Risk, and Safeguard Identification

The constitutional entity, risk and safeguard identification module extends Legal-BERT representations, identifying each clause as being related to a constitutional entity (e.g., reference to an affected person or to a public authority or to a right) or to a risk indicator (e.g., a description of automated decision making only) or to a safeguard (e.g., documented human review or data minimisation practices). It is tri-partite classification that helps differentiate between clauses that merely describe the function of a system, and clauses that have direct constitutional significance. Entities and risks are abstracted and stored as symbolic facts, along with the source clause, constitutional dimension and generated confidence score based on the Legal-BERT classification layer. Safeguard mentions are considered mitigating factors and will be taken into account when evaluating the rules, as this will be based only on identified risks, rather than all algorithmic disclosures will be treated equally when it comes to penalization.

D. Neuro-Symbolic Constitutional Rule Engine

The symbolic facts generated by the extraction module are fed into the neuro-symbolic constitutional rule engine which checks whether they satisfy a set of pre-defined constitutional rules based on expertise from judges, which are embedded in the form of principles like "legality", "proportionality", "non-discrimination", and "right to a reasoned decision". For each rule, there is a precondition on extracted facts and a risk adjustment that will be applied by the engine to a dimension's risk score when it finds a pattern that causes a concern (such as making a fully automated decision without a human review) and reduce the risk score for some pattern that leads to a safeguard (such as the fact that a documented contestability path exists). Rules are specified without relying on learning, rather than learned end-to-end, and each risk score can be fully traced to specific legal criteria, which is crucial for constitutional auditing, interpretability requirements. The engine processes all the facts extracted from a specific record, adds the risk adjustments for the dimensions as applicable and transforms the scores into the one-to-five severity scale that is utilized in the rest of the analysis.

E. Constitutional Algorithmic Risk Index (CARI)

The six risk scores that are calculated by the rule engine are turned into a single measure (in the form of a percentage) using a process called the Constitutional Algorithmic Risk Index (CARI), which allows for record-to-record comparison, dimension-to-dimension comparison, and comparison of public-sector application domains. CARI is calculated as a weighted average of the normalized scores of the dimensions, where the

weights were determined by the relative constitutional severity of the dimensions with respect to transparency, explainability, accountability, due process, fairness and privacy during expert calibration. The CARI value is used to assess the overall level of constitutional risk of the aggregate record, with three threshold values used to define moderate, high and severe risk bins during the human annotation validation process. The design enables CARI to be a diagnostic tool for specific algorithmic systems, yet also can help function as a comparative benchmark to help determine which public-sector sectors, like law enforcement or welfare administration, systematically show a higher constitutional risk that needs more concentrated regulatory interest.

$$\text{CARI} = \left[\Sigma_{\{k=1\}}^{\{6\}} \frac{w_k \cdot S_k}{\Sigma_{\{k=1\}}^{\{6\}} (w_k \cdot 5)} \right] \times 100$$

The formula used is the same as the one used for the one-to-five risk scoring for constitutional dimensions: S_k is the normalized one-to-five risk score for the given constitutional dimension k, w_k the expert-assigned severity weight for the given constitutional dimension k, and the denominator represents the maximum possible weighted risk, thus resulting in a bounded 0-100% index.

V. Experimental and Analytical Setup

A. Implementation Environment and Legal-BERT Configuration

The framework was realized in Python with the help of the HuggingFace Transformers library, and the trained checkpoint of the Legal-BERT model was the nlpaueb/legal-bert-base-uncased model.The framework was realized in Python, with the help of the HuggingFace Transformers library, and the trained checkpoint of the Legal-BERT model was the model nlpaueb/legal-bert-base-uncased. The symbolic rule engine was realized as a forward-chaining inference system and all experiments were run on a workstation that had a GPU enabled environment for the generation of embeddings in batch across the sixty-record ATRS corpus.

B. Constitutional Risk Coding and 1–5 Scoring Criteria

Each of the six dimensions was scored on an "ordinal" scale from 1 to 5, with 1 representing a minimal or well-contested risk, 3 representing the disclosure was not full or was unclear, and 5 representing the risk was severe, that is, the decision was made by computer with no human oversight or no documented contestability mechanism. A set of explicit rule definitions in the neuro-symbolic engine was used to operationalise the scoring criteria to ensure the reproducibility of the scores across the records.

C. Human Annotation and Inter-Rater Reliability

To test the automated scoring, a random sample of twenty ATRS records was manually coded by two different legal experts who are experts in public law using the same rubric (1–5) as the framework, but without knowledge of the automated output from the framework. Inter-rater reliability (Cohen's kappa) was calculated between the two human annotators, which resulted in a substantial agreement, and the agreement between the averaged scores from the human annotators and the automated scores from the framework was calculated using weighted kappa and Pearson correlation to determine criterion validity of the automated pipeline.

VI. Results and Findings

The findings from the examination of 60 ATRS records shows that while there is a moderate level of risk across most of the dimensions, the highest concentration of risk is in the transparency, explainability and accountability dimensions and the highest aggregate risk is in law-enforcement applications. Levels of concern are also progressively lower for transparency (73.33%) and human-oversight deficiency (46.67%) and there is a wide range in the CARI scores for domain-level deployments of algorithms from administrative-support (48.60%) to law-enforcement (82.40%).

A. ATRS Algorithmic System and Public-Authority Distribution

The sixty ATRS records are distributed across six domains: law enforcement, welfare, immigration, fraud detection, recruitment, and administrative support, with law-enforcement and welfare records accounting for more than half of the total, and reflecting the high stakes and rights-affecting nature of public functions where algorithms are being used.

B. Constitutional Risk Distribution Across ATRS Records

Table 1 and Figure 2 show the distribution of the constitutional risk in all the evaluated ATRS records. The most important dimension is transparency at 73.33% and mean risk score of 4.18 followed very closely by explainability at 70.00% and accountability at 68.33%. Together, these three dimensions rank on top, meaning that government algorithmic systems have the most widespread constitutional shortcomings in the form of

problems with reason giving and disclosure. Proportionality, contestability and human-oversight deficiency see moderate degrees of concern; this is in comparison to the other procedural protections, which are found at progressively lower degrees of concern: due process, fairness, and privacy in order.

Table 1. Constitutional Risk Distribution across ATRS Algorithmic Systems

Constitutional Parameter	Concern (%)	Mean Risk Score (1–5)	Risk Level	Rank
Transparency	73.33	4.18	High	1
Explainability	70.00	4.06	High	2
Accountability	68.33	3.94	High	3
Due Process	65.00	3.82	High	4
Fairness/Equality	60.00	3.67	High	5
Privacy	56.67	3.51	Moderate–High	6
Proportionality	51.67	3.34	Moderate	7
Contestability	48.33	3.21	Moderate	8
Human-Oversight Deficiency	46.67	3.08	Moderate	9

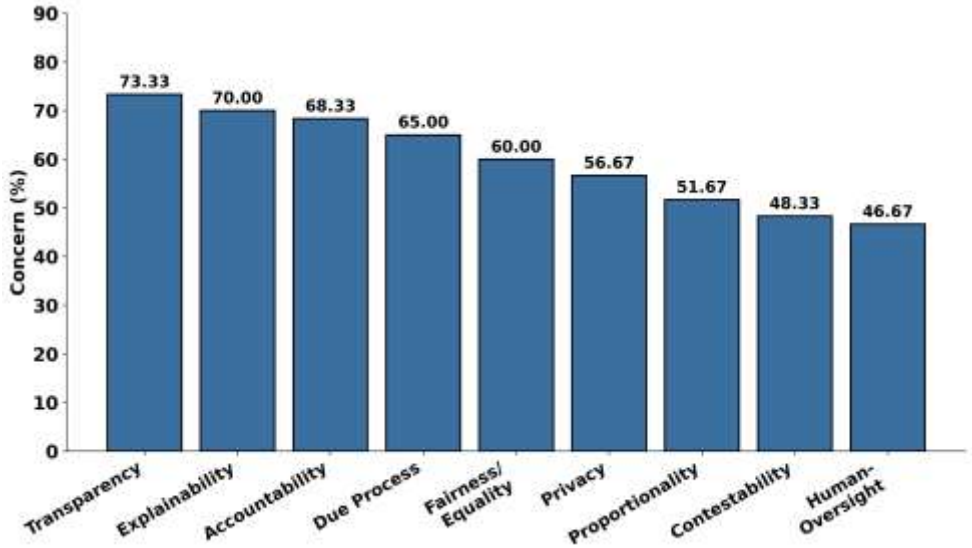

Figure 2. Concern percentages across nine constitutional risk parameters derived from ATRS record analysis.

As illustrated in Figure 2, the dimension of disclosure-related risk seems to be the most prevalent in identified risk, from a concern level of transparency to human-oversight deficiency.

C. Transparency, Fairness, Privacy, Due Process, and Accountability Analysis

Transparency, explainability and accountability have the highest constitutional concern levels at 73.33%, 70.00% and 68.33% respectively and the lowest concern levels are those of fairness (60.00%) and privacy (56.67%), though concern levels are still reasonably high in these dimensions, which addresses discriminatory and data-protection risk somewhat more consistently than disclosure-related risk.

Table 2. Severity Analysis of Major Constitutional Risks

Constitutional Dimension	Low (%)	Moderate (%)	High (%)	Severe (%)	High + Severe (%)
Transparency	10.00	16.67	43.33	30.00	73.33

Explainability	11.67	18.33	41.67	28.33	70.00
Accountability	13.33	18.34	40.00	28.33	68.33
Due Process	15.00	20.00	38.33	26.67	65.00
Fairness/Equality	16.67	23.33	36.67	23.33	60.00
Privacy	20.00	23.33	35.00	21.67	56.67

Table 2 and Figure 3 further expand transparency, explainability and accountability, due process, fairness and privacy into severity bands of low, moderate, high and severe. Transparency is again the highest proportion with a 'high-severe' rating (73.33% of records) and a proportion of severe records (30.00%) meaning opacity is not just common, but often serious. Similarly, with explainability, 28.33 % of the classifications for the severe classes, indicating that there is a structural weakness with regard to an inadequate explanation of the algorithm, which is not a problem that only affects a small number of algorithmic systems.

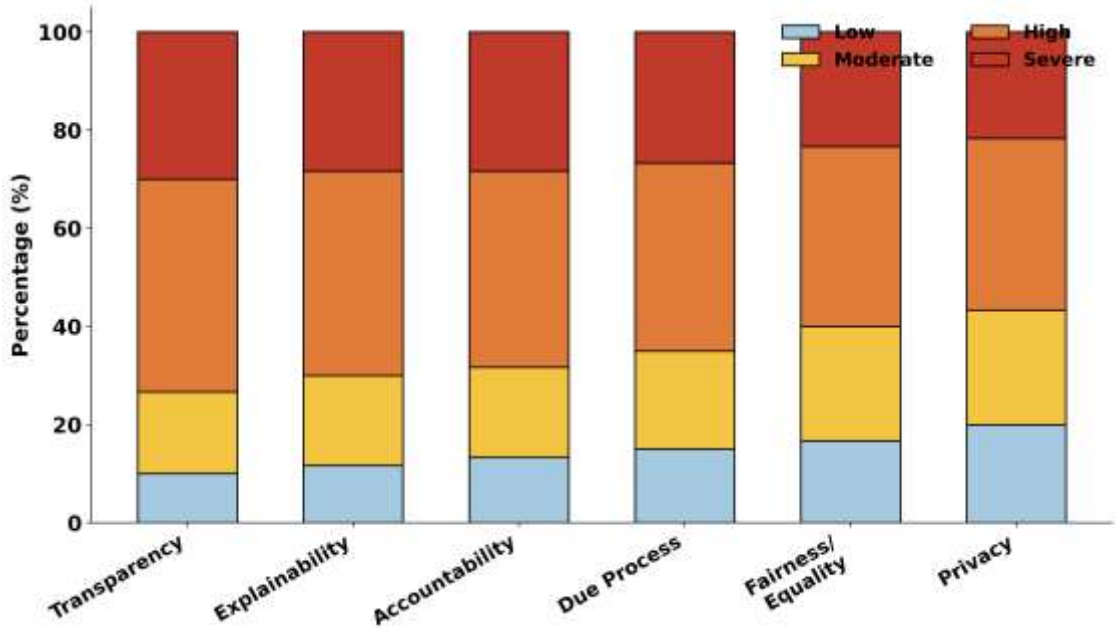

Figure 3. Severity-band decomposition (low, moderate, high, severe) for six major constitutional risk dimensions.

Stacked severity proportions in Figure 3 clearly indicate that the dimensions of transparency and explainability have by far the highest severity risk shares in both the high and severe risk areas.

D. Human Oversight, Contestability, and Safeguard Analysis

Human-oversight deficiency and contestability rank the least at 46.67% and 48.33% respectively, but remain in the moderate risk band, indicating that whilst many systems claim to have a human-oversight element, the level of such oversight and its enforceability differs significantly and protection is not always constitutionally complete.

E. CARI-Based Comparative Constitutional Risk Analysis

The overall aggregate CARI component scores for all public-sector domains is 77.27%, and for the law-enforcement sector is 82.40%, demonstrating that constitutional risk is not evenly distributed across the different sectors of the public sector and that a one-size-fits-all governance approach is not suitable.

Table 3. CARI-Based Risk across Algorithmic State Application Domains

Application Domain	Transparency (%)	Fairness (%)	Privacy (%)	Due Process (%)	Mean CARI (%)
Law Enforcement	83.33	75.00	66.67	75.00	82.40

Welfare/Public Benefits	75.00	75.00	50.00	83.33	78.60
Immigration/Border	80.00	70.00	70.00	70.00	77.80
Fraud/Risk Detection	70.00	50.00	60.00	60.00	68.40
Public Recruitment	62.50	75.00	37.50	50.00	64.20
Administrative Support	50.00	37.50	37.50	37.50	48.60

Table 3 and Figure 4 present the percentages of each application domain that are constitutional transparent, constitutional fair, constitutional privacy, and constitutional due-process as well as application domain mean CARI scores for constitutional risk across the six application domains. The highest mean CARI (82.40%) is found within law-enforcement applications, which is in severe category, as a result of high level of transparency and concerns of due process. In comparison, the mean CARI for welfare and immigration applications are relatively high at 78.60% and 77.80% respectively, both of which are considered high-risk applications, given their comparatively high stakes and discretion for decisions, and administrative-support applications are comparatively low at 48.60%, reflecting the lower-stakes, low-discretion nature of the decision-making context.

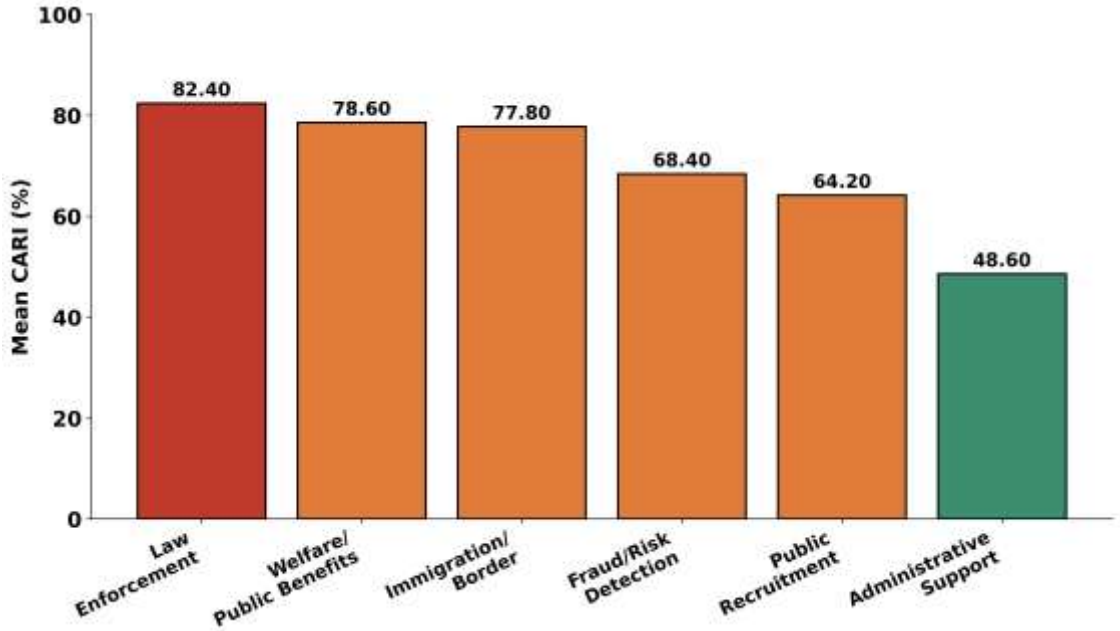

Figure 4. Mean Constitutional Algorithmic Risk Index (CARI) compared across six public-sector application domains.

Figure 4 indicates the mean CARI for each domain with law enforcement being the most and administrative support being the least.

VII. Discussion and Constitutional Implications

A. Algorithmic State Power and Constitutional Rights

Algorithmic state power "adds a layer of statistics between people and their rights and liberties in the constitution. Overall, the issues raised in the ATRS highlights that the use of algorithms in many government systems is currently opaque, lacking in explanation, and fails to provide a basis for review of government action, which directly impacts the constitutional right to have government power exercised in known, thoughtful, and reviewable ways. Even when citizens have the right to be treated the same way as others before the state, and

to not be denied a benefit without good reason, the substantive content of such rights is made hard to put into practice without an automated system identifying the citizens as high risk. The serious CARI classification in the law-enforcement domain is especially alarming given its implications for liberty interests, search and seizure protections and the presumption of innocence which form the basis of criminal-justice systems in constitutional democracies. These findings all point to the possibility that, if left unchecked, algorithmic state power will work the Constitution's protections away from individuals and towards the technical and institutional players behind the design, procurement and operation of government AI systems – a transformation that flies in the face of the Constitution's fundamental principle: that public power should be subject to the people.

B. Explainability, Accountability, and Rule-of-Law Implications

Strong associations between explainability deficits and accountability weaknesses are observed for each ATRS record, highlighting a structural rule-of-law issue: accountability, such as administrative appeals, judicial review, etc., requires basic requirements of accountability, namely, intelligible reasons, and where there is a lack of explainability, so goes the practical power of accountability. This interdependency means there is a need to implement piecemeal reforms that weave in and out of just one aspect, such as requiring disclosure but not interpretability.It implies that piecemeal reforms such as those requiring disclosure only without requiring interpretability will not meaningfully improve adherence to the rule of law. The rule-of-law tradition holds that the exercise of state power must be guided by rules, which are known to the public, and in advance, clear and predictable, and that they must apply uniformly and consistently: machine-learning models are not uniformly and consistently clear, predictable and known to the public, and, even if they are disclosed, they are only imperfectly these. The accountability of algorithmic state power thus needs to be addressed at the same time at three levels explainability, documentation, and institutional-review and not at one or two. The moderate risk of contestability also suggests that, even if there is a process in place to appeal decisions, it remains unclear how likely it is that they are made available and if they are effective. This has implications for the right to a fair hearing, arising from real decisions that are likely to have a significant impact on the life of an individual, before such a process.

C. Constitutional Safeguards for AI-Driven State Decisions

The CNSRA framework and CARI metric identify a number of specific constitutional protections that can help reduce the risks identified in the use of AI in the state decision-making process. Algorithmic impact statements, which are required and needed to be written in plain language, and to explain the reasoning behind algorithms instead of technical documentation, could directly tackle the transparency/ explainability gaps that were found to be the top risk dimensions. Requiring high-stakes domains like law enforcement and welfare to have some human in the loop review, and provide documented and auditable override, would increase accountability and due-process compliance over disclosure-only. Founding independent algorithmic-audit systems that are given the power to apply a framework like CARI (constitutional audit) regularly will institutionalize the constitutional oversight, not just with voluntary transparency reporting. Proportionality assessments, in which public authorities have to prove that the automated decision is necessary, and that there is no alternative less rights-restrictive, would also make the deployment of algorithms correspond to other constitutional balancing principles already known to courts and administrative authorities.

D. Policy Implications, Limitations, and Generalizability

Policy implications of the above findings are that policy should aim to ensure a differentiated intensity of regulation depending on the constitutional risk faced in the particular domain, with the highest level of regulation applied to the law-enforcement, welfare and immigration applications where CARI scores are severe or high on average. Although there are several limitations that need to be taken into account when making the conclusions, the sixty-record ATRS sample, although structurally diverse, is limited to a single national jurisdiction and disclosure regime and immediate generalizability is therefore limited to jurisdictions with differing disclosure obligations or legal traditions. Although the rule base is expert-calibrated, it does reflect the interpretation of constitutional severity that could be interpreted differently by other legal scholars, and is different from the interpretation of constitutional severity of other constitutional systems. The generalizability of CARI could be further enhanced through validation in other jurisdictions, with larger numbers of records and in alternative constitutional contexts, in the future.

VIII. Conclusion and Future Work

In this research, we present a framework, named Constitutional Neuro-Symbolic Risk Assessment, that combines the explicit symbolic reasoning and the Legal-BERT semantic representation to quantify the

constitutional risk in government algorithmic systems and present it as an algorithmic risk assessment tool. The model applied to 60 ATRS records found that, on average, CARI scores were 77.27% (with a range of -0.43 to 0.96) and law-enforcement applications had a median score of 82.40% (range -0.43 to 0.96). The findings in this report show that as disclosure and governance of algorithmic state power is currently shaped, it often does not meet the fundamental expectations of the rule of law that are crucial for reasoned, contestable and accountable decision-making. The proposed structure brings a replicable and explainable auditing approach from the legal-language intelligence and constitutional law fields to provide a tool for regulators, courts and public authorities to systematically monitor algorithms. The results highlight the need to establish constitutional values directly in the design of systems, not just in a post hoc review of algorithmic systems, when it comes to protecting fundamental rights in an algorithmic age. The recommendations are the implementation of impact statements with mandatory plain language, as well as a human touch applied at a domain level, independent regular audits and proportionality assessments applied to the higher risk domains. Future research could expand the validation of CARI to other jurisdictions, integrate the tracking of algorithmic systems' evolution over time, and explore the incorporation of existing legal-compliance verification tools, and develop a more transparent, accountable and democratically legitimate approach to algorithmic governance.

References

[1] I. Bar-Siman-Tov, "Legislatures and legislation in the age of artificial intelligence," The Theory and Practice of Legislation, vol. 13, no. 3, pp. 267–291, 2025.
[2] M. Araszkiewicz and M. Florczak-Wątor, "AI and the principles of proper legislation: enhancing quality, understandability, and consistency in legal texts," The Theory and Practice of Legislation, vol. 13, no. 3, pp. 353–382, 2025.
[3] V. Nizov, "The Artificial Intelligence Influence on Structure of Power: Long-Term Transformation," Legal Issues in the Digital Age, vol. 6, no. 2, pp. 183–212, 2025.
[4] M. Kuziemski and G. Misuraca, "AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings," Telecommunications Policy, vol. 44, no. 6, p. 101976, 2020.
[5] Y. Rymon, "Of the people, by the algorithm: how AI transforms the role of democratic representatives?," AI & Society, vol. 41, pp. 5997–6012, 2026.
[6] G. Frosio and C. Geiger, "Taking fundamental rights seriously in the Digital Services Act's platform liability regime," European Law Journal, vol. 29, no. 1–2, pp. 31–77, 2023.
[7] A. Ovadya, "Reimagining Democracy for AI," Journal of Democracy, vol. 34, no. 4, pp. 162–170, 2023.
[8] R. H. Weber, "Artificial Intelligence ante portas: Reactions of Law," J, vol. 4, no. 3, pp. 486–499, 2021.
[9] I. Uhumuavbi, "An Adaptive Conceptualisation of Artificial Intelligence and the Law, Regulation and Ethics," Laws, vol. 14, no. 2, p. 19, 2025.
[10] J. He and Z. Zhang, "Algorithm Power and Legal Boundaries: Rights Conflicts and Governance Responses in the Era of Artificial Intelligence," Laws, vol. 14, no. 4, p. 54, 2025.
[11] G. F. Lendvai and G. Gosztonyi, "Algorithmic Bias as a Core Legal Dilemma in the Age of Artificial Intelligence: Conceptual Basis and the Current State of Regulation," Laws, vol. 14, no. 3, p. 41, 2025.
[12] N. M. Pawar, R. M. Sairise, P. Karemore, A. Jadhav, D. Turaev, and S. N. Ajani, "ML-based verification systems using discrete mathematical models," Journal of Discrete Mathematical Sciences and Cryptography, vol. 29, no. 2-B, pp. 971–978, 2026, doi: 10.47974/JDMSC-2548.
[13] G. N. Mukhamadieva, A. K. Zhanibekov, N. M. Apsimet, and Y. T. Alimkulov, "Integration of Artificial Intelligence into Criminal Procedure Law and Practice in Kazakhstan," Laws, vol. 14, no. 6, p. 98, 2025.
[14] N. Manos, E. Technitis, and A. Sykiotou, "The Use of Artificial Intelligence in the Administration of Justice: Suggested Framework of Ethical Principles and Reasoning of Judges in the Use of Intelligent Systems," Laws, vol. 15, no. 2, p. 20, 2026.
[15] P. Dumouchel, "AI and Regulations," AI, vol. 4, no. 4, pp. 1023–1035, 2023.
[16] P. Casanovas, L. de Koker, and M. Hashmi, "Law, Socio-Legal Governance, the Internet of Things, and Industry 4.0: A Middle-Out/Inside-Out Approach," J, vol. 5, no. 1, pp. 64–91, 2022.
[17] L. Hogan and M. Lasek-Markey, "Towards a Human Rights-Based Approach to Ethical AI Governance in Europe," Philosophies, vol. 9, no. 6, p. 181, 2024.
[18] S. V. Shinde, T. Shitole, S. Mundhe, S. Kalamkar, and V. Rajput, "Exploration Of Recommendation System for Service Discovery," International Journal on Advanced Computer Theory and Engineering, vol. 14, no. 1, pp. 11–15, 2025.