

Carlos Ramírez Castañeda

Use of Artificial Intelligence by Public Authorities for the Detection and Prevention of Cyberterrorism

Abstract

The use of artificial intelligence by public authorities to detect and prevent risks associated with cyberterrorism raises a central tension between institutional effectiveness and constitutional limits. These tools make it possible to process large volumes of information, identify anomalies, and protect critical infrastructure, yet their legitimacy cannot be measured solely by their operational usefulness. In a democratic state governed by the rule of law, algorithmic prevention requires a sufficient legal basis, a legitimate aim, proportionality, personal data protection, non-discrimination, meaningful human oversight, traceability, and judicial control. This article develops a legal-documentary and comparative analysis of the use of artificial intelligence systems in early detection, risk assessment, and cyberterrorism prevention. It examines the Budapest Convention, the UNESCO and OECD recommendations, the NIST AI Risk Management Framework and Regulation (EU) 2024/1689, together with the experience of ENISA and CISA, the United Nations frameworks on artificial intelligence and counter-terrorism, and three judicial decisions that have subjected algorithmic risk systems to review. It concludes that artificial intelligence can only be legitimately employed when it is embedded in a guarantee-based architecture of legality, auditing, accountability, and effective review.

Keywords: artificial intelligence; cyberterrorism; public authorities; fundamental rights; cybersecurity; algorithmic governance; critical infrastructure

Carlos Ramírez Castañeda, PhD in Public Administration and Public Policy, Postdoctoral Fellow in Law. National Open and Distance University of Mexico (UnADM).

Received
05/18/2026

To cite this article: Ramírez, C. (2026), Use of Artificial Intelligence by Public Authorities for the Detection and Prevention of Cyberterrorism, *Revista Internacional de Estudios sobre Terrorismo*, issue 18:7-29.

Accepted
08/24/2026

1. Introduction

Artificial intelligence has become one of the most relevant tools in the digital transformation of the public sector. Its incorporation into state functions has generated expectations of efficiency, anticipation, and optimization in areas as diverse as health, administration, justice, security, and the management of critical infrastructures. In the field of cybersecurity, these expectations intensify, since AI offers capacities for mass processing, anomaly detection, incident classification, and decision-support in scenarios characterized by speed, complexity, and uncertainty.

When these capacities are oriented toward the detection and prevention of cyberterrorism, the debate acquires a singular legal density. The question is no longer solely whether the technology can improve the State's response, but whether its use by public authorities is compatible with the principle of legality, fundamental rights, and the material limits of preventive power. In this context, AI stands at the intersection of two constitutional requirements: on the one hand, the State's obligation to protect the population and ensure the continuity of essential services; on the other, the necessity that every public intervention remain subject to democratic control and to the guarantees proper to the rule of law.

The difficulty increases owing to the conceptual instability of cyberterrorism itself. It is a category frequently employed in an expansive manner, mixing attacks on critical infrastructure, sabotage, online extremist propaganda, coordinated campaigns of digital intimidation, and other manifestations of technological criminality. As Macdonald et al. (2022) note, the academic community has not reached a definitive consensus on the constitutive elements of cyberterrorism, which hinders its normative treatment and its distinction from related figures such as hacktivism, digital espionage, or cyberwarfare. This conceptual breadth is legally problematic, for it may lead both to the banalization of the term and to excessive state responses. For this reason, a serious analysis of the use of AI by authorities must begin with a careful delimitation of the preventive object.

The study is situated within a doctrinal debate that has gained density in recent years. The literature on algorithmic governance has shown that automated risk-management systems are not mere auxiliary instruments but devices that reconfigure the very manner in which public power defines, measures, and anticipates danger (Yeung, 2018). In parallel, European and Ibero-American public law has begun to discuss the legal nature of the algorithms employed by the administration and the standard of legal authorization that their use requires (Boix Palop, 2020; Cotino Hueso, 2023). To this is added specific institutional production on artificial intelligence and counter-terrorism, driven by the United Nations, which warns of the risks of opacity, bias, and over-surveillance associated with these applications (UNOCT and UNICRI, 2021; OHCHR, 2025). The present work dialogues with that literature and seeks to contribute a reading situated in the Mexican legal order.

The purpose of the present article is to examine under what conditions the use of artificial intelligence by authorities to detect and prevent cyberterrorism may be considered legally legitimate.

The thesis advanced is that the validity of these tools does not derive from their technical utility alone. Their legitimacy depends on a rights-protective architecture based on legality, legitimate purpose, proportionality, data protection, meaningful human oversight, traceability, independent audit, and the existence of effective mechanisms of review and accountability.

2. Methodology

The research is qualitative, legal-documentary, and analytical-comparative. It is based on the review of specialized academic literature on cyberterrorism, artificial intelligence, public function, algorithmic governance, fundamental rights, and technological risk management. Likewise, relevant legal instruments and international frameworks of reference are examined in order to assess the use of AI in sensitive contexts of state prevention.

At the national level, the Political Constitution of the United Mexican States and the Federal Penal Code are taken as referents, with the aim of situating the discussion within the framework of the principles of legality, due process, human rights, and the prosecution of serious threats. At the international and comparative level, the following are reviewed: the United Nations Convention against Transnational Organized Crime, the Council of Europe Convention on Cybercrime, the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD Council Recommendation on Artificial Intelligence, the NIST AI Risk Management Framework, and Regulation (EU) 2024/1689 of the European Parliament and of the Council.

The comparative approach does not seek to transfer foreign solutions mechanically to the Mexican context. Rather, it aims to identify functional principles and normative standards that make it possible to evaluate the lawfulness of the state use of algorithmic systems in preventive tasks. To this end, the analysis is organized around five variables: legitimate purpose, sufficient normative basis, intensity of impact on fundamental rights, level of autonomy of the system, and existence of ex ante, concurrent, and ex post controls.

Alongside the normative analysis, the work incorporates three judicial decisions that have subjected to review algorithmic risk-assessment systems deployed by public authorities. Their selection responds to three criteria: that the system examined had been operated by a public power for preventive or security purposes; that the decision had expressly addressed the sufficiency of the legal authorization or the guarantees against automated processing; and that the pronouncement had come from a judicial body whose doctrine is transferable, as an argumentative standard, to the constitutional examination proper to the Mexican legal order. These decisions do not operate as binding precedents, but as cases that make it possible to observe how the five indicated variables are applied in real litigation scenarios.

From an epistemological standpoint, the research adopts a critical legal approach that recognizes that law is not neutral in the face of technology. As Danaher et al. (2017) point out, legal institutions

tend to lag behind technological development, and that gap may generate normative voids with direct consequences for the protection of rights. For this reason, the analysis is not limited to describing the current state of regulation but evaluates its sufficiency and formulates proposals for improvement oriented toward guaranteeing the legitimate exercise of the State's preventive power.

3. Cyberterrorism and Algorithmic Prevention: Necessary Clarifications

3.1. *Definition and Delimitation of Cyberterrorism*

The category of cyberterrorism requires careful delimitation. Not every high-impact computer incident can be classified as such, nor does every criminal activity in cyberspace entail the purpose of collective intimidation or institutional destabilization proper to terrorism. The decisive criterion is found not in the digital medium employed, but in the purpose pursued and in the nature of the interests affected.

Specialized literature has insisted that cyberterrorism involves the use of digital means with the objective of instilling fear, coercing authorities, or altering essential functions of the State or of social life (Denning, 2000; Conway, 2017; Mayer Lux, 2018). Denning (2000) defined cyberterrorism as unlawful attacks and threats of attack against computers, networks, and the information stored therein when carried out to intimidate or coerce a government or its population in the pursuit of political or social objectives. This definition, pioneering in the field, continues to be an obligatory reference in the specialized literature, although it has been the object of subsequent refinements.

Macdonald et al. (2022) conducted a survey of specialist terrorism researchers and found that there is broad consensus on the need to distinguish cyberterrorism from ordinary cybercrime and from hacktivism on the basis of the intentional element: while the first seeks to generate terror or political destabilization, the second pursues economic gains or the temporary interruption of services, and the third has protest motivations without necessarily producing damage of the magnitude proper to terrorism. This distinction has first-order legal implications, for it determines the applicable regime, the procedural guarantees required, and the proportionality of preventive measures.

For their part, Venkatachary et al. (2024) analyze the specific challenges that cyberterrorism poses for critical infrastructures in the energy sector, identifying differentiated attack vectors, applicable cybersecurity frameworks, and economic dimensions of the problem. Their analysis confirms that attacks on essential infrastructures may produce cascading effects with consequences for public security, health, and the continuity of the State, which justifies a high-intensity preventive response, albeit subject to strict legal conditions.

The UNODC, in its educational module on cybercrime (n.d.), distinguishes among hacktivism, terrorism, espionage, disinformation, and cyberwarfare as analytically separate categories, each with its own characteristics regarding actors, motivations, targets, and legal consequences. This taxonomy

is useful because it demonstrates that algorithmic prevention cannot operate with undifferentiated categories: the system must be designed to detect legally typified conduct, not to track generically “suspicious activities” without precise normative contours.

3.2. Algorithmic Prevention: Technical and Legal Characterization

In parallel, algorithmic prevention may be understood as the use of computational systems capable of identifying patterns, inferring risks, classifying events, or prioritizing alerts in order to orient institutional response. In public cybersecurity contexts, such prevention may manifest itself in anomaly-detection systems in networks, incident-correlation tools, automated analysis of indicators of compromise, or natural-language processing of open sources. From a technical perspective, AI systems applied to cybersecurity may be classified into various categories according to their function: machine-learning-based intrusion-detection systems (ML-IDS), threat-intelligence platforms (TIP), user-and-entity-behavior-analytics tools (UEBA), AI-enhanced security-information-and-event-management systems (SIEM), and natural-language models for the analysis of content in open sources (OSINT). Each type of system raises distinct legal constraints in terms of the data processed, the level of autonomy, and the potential impact on rights (European Union Agency for Cybersecurity, 2023; National Institute of Standards and Technology, 2023). It is also necessary to specify what type of algorithmic regulation is at stake. Yeung (2018) proposes distinguishing between reactive systems, which respond to already verified behaviors, and preventive or anticipatory systems, which act upon inferences about future conduct, and warns that the latter concentrate the most serious problems of legitimacy, since they intervene before a legally relevant fact exists. The detection of cyberterrorism is situated precisely within this second group. Hence the examination cannot be limited to the statistical reliability of the model: what is under discussion is the lawfulness of anticipating legal consequences on the basis of a calculated probability. The legal complexity increases, however, when these systems cease to be simple technical supports and begin to influence institutional decisions concerning surveillance, risk prioritization, or the opening of lines of investigation. The difference between operational support and the structuring of preventive judgment is central. In the first case, the system may serve to strengthen the resilience of networks or infrastructures. In the second, AI begins to affect rights, privacy expectations, monitoring patterns, or administrative and criminal processes directly. Regulation (EU) 2024/1689, known as the Artificial Intelligence Act, establishes precisely this distinction as a criterion for risk classification. AI systems used by authorities for law enforcement, risk assessment, or the identification of suspects are considered high-risk and are subject to reinforced obligations of transparency, documentation, human oversight, and conformity assessment prior to deployment. This normative categorization confirms that European law already recognizes that not all uses of AI in the public sphere are equivalent and that regulatory intensity must be graduated according to the potential impact on fundamental rights. Consequently, the legal discussion should not focus solely on the technical sophistication of the model, but on the institutional place it occupies within state action.

4. Legal Risks of the Use of AI by Public Authorities

4.1. *Insufficient Normative Basis*

The first legal risk is the insufficiency of the normative basis. No public authority can justify the intensive use of AI in preventive functions on the basis of generic clauses of security or administrative modernization. The principle of legality requires a clear authorization regarding purposes, competent authorities, categories of data, temporal limits, activation scenarios, and oversight mechanisms. Without a sufficient normative basis, the technological tool ends up substituting for the norm and blurring the guarantee of foreseeability.

Administrative-law doctrine has offered a path for specifying this requirement. Boix Palop (2020) maintains that the algorithms employed by the administration to adopt decisions fulfill a function materially equivalent to that of a regulation, insofar as they regulate and predetermine the action of public power and should therefore be subject to the guarantees of publicity, participation, and control already provided for regulatory norms. Cotino Hueso (2023), examining biometric-identification systems, emphasizes that the overlap between artificial-intelligence regulation and data-protection regulation does not dispense with a specific and high-quality legal authorization, and that the standards demanded by various European constitutional courts are rarely satisfied in administrative practice. Both approaches converge on a point relevant to this work: the reservation of law is not exhausted by the existence of a norm that mentions the technology; rather, it demands normative density concerning its design, its limits, and its controls.

The European Court of Human Rights has developed in its case-law the doctrine of the “quality of the law” as a condition for interferences with fundamental rights to be compatible with the European Convention. According to this doctrine, the enabling law must not only exist but be accessible, foreseeable, and endowed with adequate safeguards against abuse. Although this case-law is not directly binding on Mexico, it offers a useful standard for evaluating the sufficiency of normative authorizations in matters of surveillance and algorithmic prevention.

In the Latin American context, the Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights has alerted to the risks of digital surveillance without clear legal authorization, noting that the interception of communications and the mass processing of data by the State require specific normative frameworks that establish with precision the scenarios, the competent authorities, and the control mechanisms. Although its pronouncements refer primarily to the surveillance of communications, the principles they develop are fully applicable to algorithmic systems for the detection of cyberterrorism.

4.2. *Protection of Personal Data*

A second risk lies in the protection of personal data. The preventive logic tends to favor the accumulation of information under the premise that more data improve detection capacity. That assumption, however, must be subjected to the principles of minimization, specific purpose, proportionality, accuracy, and security. Mass and indiscriminate collection is not only legally problematic but also technically questionable, for it may increase informational noise and produce erroneous or excessive inferences.

In Mexico, the General Law on the Protection of Personal Data Held by Obligated Subjects establishes guiding principles applicable to the processing of data by public authorities, including lawfulness, purpose, loyalty, proportionality, and accountability. Nevertheless, the law contains no specific provisions on the processing of data in algorithmic systems of public security, thereby maintaining a zone of normative insufficiency that future regulation will have to address.

UNESCO (2021), in its Recommendation on the Ethics of AI, pays special attention to the protection of privacy and personal data as a necessary condition for the ethical development and deployment of AI systems. The Recommendation establishes that AI systems must not be used to infringe the right to privacy, which includes control over personal data and protection against arbitrary or unlawful surveillance. These principles are directly applicable to the design of algorithmic systems for the detection of cyberterrorism by public authorities.

4.3. *Bias, Discrimination, and Epistemic Validity*

In the third place, the problem of bias and discrimination must be considered. Algorithmic systems learn from data, design decisions, and institutional contexts. If those elements are marked by asymmetries, errors, or logics of over-surveillance, the tool may reproduce or amplify such distortions. In the field of cyberterrorism prevention this is especially serious, since an erroneous classification may activate enhanced surveillance, affect reputations, or generate disproportionate treatments before effective judicial control exists.

Specialized literature has extensively documented the risks of bias in AI systems used in security contexts. Studies such as those of Obermeyer et al. (2019) have demonstrated that algorithms employed in public decision-making may encode and amplify racial, socioeconomic, or gender biases present in training data. In the specific domain of security, biases may translate into the over-surveillance of particular communities, generating a disproportionate impact on persons and groups already in situations of vulnerability. This dynamic is incompatible with the principle of equality and non-discrimination, enshrined in Article 1 of the Political Constitution of the United Mexican States and Article 24 of the American Convention on Human Rights.

The NIST AI Risk Management Framework (2023) addresses the problem of bias as a systemic risk that must be managed at all stages of the system life-cycle: design, development, validation, deployment, and continuous monitoring. The framework proposes mechanisms for fairness evaluation, adversarial testing, and feedback mechanisms to detect and correct biases, underscoring that bias management is not a one-time event but a permanent process. This perspective is especially relevant for cyberterrorism-detection systems, where the dynamics of threats change constantly and training data may rapidly become obsolete or biased.

4.4. Opacity and the Demand for Explainability

A fourth risk is opacity. When the authority cannot sufficiently explain how the system operated, which data were relevant, or how the algorithmic recommendation was integrated into the final decision, the requirement of reasoned motivation is weakened and judicial control becomes precarious. For this reason, traceability and explainability are not desirable attributes in the abstract, but conditions of lawfulness.

The OECD (2023), in its revised Recommendation on AI, establishes that AI systems must be transparent and that operators must be able to explain their decisions and predictions to the extent reasonably possible. It further notes that AI agents should be able to provide sufficient information to help stakeholders understand the capabilities and limitations of the systems and to respond to the outcomes they produce. These explainability requirements are especially relevant in the context of the state use of AI, where decisions taken on the basis of algorithmic systems may affect fundamental rights and must be capable of effective judicial review.

Regulation (EU) 2024/1689 operationalizes these requirements through concrete obligations for high-risk systems, including the keeping of automatic logs, detailed technical documentation, the provision of instructions for use, and the obligation to design the system so that its outputs are interpretable by the natural persons responsible for its oversight. These obligations not only have instrumental value for judicial control but also contribute to building the institutional trust necessary for the legitimate deployment of AI in sensitive contexts.

Doctrine has nevertheless pointed to the limits of this model. Veale and Zuiderveen Borgesius (2021) warn that a substantial part of the transparency obligations of the European regime rests on mechanisms of self-assessment of conformity, so that compliance remains in the hands of the provider itself or of the deploying organization, with limited external oversight. This observation is especially pertinent in the security domain, where requirements of secrecy may be invoked to restrict access to technical documentation. If the explanation is reduced to an internal file inaccessible to the affected person and to the judge, the guarantee becomes nominal.

4.5. Public Responsibility and the Risk of Technocracy

Finally, there exists the risk of a weakening of public responsibility. The administration cannot shield itself behind the technical complexity of the system or the prestige of the provider in order to dilute its responsibility. If the technology influences a state action, legal imputation continues to correspond to the authority. Meaningful human oversight, in this sense, is also a requirement of democratic responsibility.

The risk of technocracy manifests itself when the technical complexity of algorithmic systems generates an informational asymmetry between technological operators and political and judicial decision-makers. This asymmetry may produce what doctrine terms “algorithmic deference”: the tendency of officials to accept without questioning the recommendations of automated systems, effectively delegating the decision to the algorithm. In contexts of cybersecurity and terrorism prevention, where operational urgency is high, this risk intensifies.

The OECD Recommendation on AI (2023) addresses this problem by requiring that AI systems be designed in such a way as to permit oversight, replacement, or reversal by human operators. UNESCO (2021) goes further by establishing that decisions that significantly affect the rights and freedoms of persons must be capable of being reviewed and, if necessary, corrected by human beings.

5. International and Comparative Referents

5.1. United Nations Convention against Transnational Organized Crime

The United Nations Convention against Transnational Organized Crime does not directly regulate artificial intelligence or cyberterrorism, yet it provides a valuable architecture for international cooperation against complex and cross-border phenomena. Its principal teaching is that the state response to organized risks cannot rest solely on technical tools, but on institutional coordination, information exchange, and capacity strengthening. The Convention establishes, in its Article 29, the obligation of States Parties to adopt, develop, or improve training programmes for their personnel, including training in investigative techniques compatible with human rights. This provision is relevant to the use of AI in the prevention of cyberterrorism, since it implies that any technological tool must be accompanied by adequate training that guarantees its use within the parameters of the rule of law.

5.2. Budapest Convention on Cybercrime

The Budapest Convention, for its part, offers an essential framework in matters of criminal cooperation and the preservation of digital evidence. Its relevance in this study lies in recalling that technological detection does not replace the lawfulness of the investigation. AI may contribute to identifying early signals or threat patterns, but any subsequent action must remain integrated into

procedures compatible with due process and with the rules governing the obtaining and preservation of evidence. The Budapest Convention, in its Chapter II, establishes specific provisions on procedural measures that States must adopt for the investigation of computer-related offences, including the expedited preservation of stored computer data, the expedited disclosure of preserved traffic data, and the interception of content data. These provisions must be understood as a minimum floor of guarantees that AI systems used in the detection of cyberterrorism cannot ignore.

5.3. UNESCO and OECD Ethical Frameworks

The UNESCO Recommendation on the Ethics of Artificial Intelligence and the OECD Recommendation on Artificial Intelligence provide principles directly useful for evaluating the state use of AI: respect for human rights, privacy, transparency, explainability, accountability, security, robustness, and human oversight. Both instruments show that the public sector, especially in sensitive domains such as security and justice, must operate with more demanding standards than purely private or experimental contexts.

UNESCO (2021) introduces the concept of “ethical impact assessment” as a prospective tool for identifying the possible effects of AI systems on human rights prior to their deployment. This tool is especially relevant for cyberterrorism-detection systems, given that the anticipation of effects on privacy, equality, and due process may contribute to designing systems more respectful of rights from the outset, incorporating what doctrine terms “privacy by design”.

The OECD (2023), for its part, has updated its Recommendation on AI to reinforce the principles of accountability and responsibility, noting that AI actors should be accountable for the proper implementation of AI-related principles. In the public-sector domain, this responsibility acquires a constitutional dimension, since the State does not act as merely another private agent, but as the holder of public power whose action is subject to the principles of legality, effectiveness, and service to the general interest.

5.4. NIST AI Risk Management Framework

The NIST AI Risk Management Framework adds an especially valuable operational approach: the continuous management of risk. The governance of AI should not be conceived as a one-time authorization, but as a permanent process of identification, assessment, mitigation, monitoring, and review. In preventive contexts, where risk and uncertainty change constantly, this logic is indispensable.

The NIST AI RMF 1.0 (2023) structures risk management into four principal functions: Govern, Map, Measure, and Manage. The Govern function establishes the policies, processes, procedures, and practices for the governance of AI risk. The Map function contextualizes the identified risks according to purpose and stakeholders. The Measure function analyzes, assesses, and monitors risks.

And the Manage function prioritizes risk responses and applies appropriate measures. This framework is directly applicable to the design of cyberterrorism-detection systems by authorities.

5.5. Regulation (EU) 2024/1689 and the Experience of ENISA

Regulation (EU) 2024/1689 confirms that the greater the sensitivity of the use environment and the greater the possibility of impact on fundamental rights, the denser the regulatory obligations must be. Its comparative value for the Mexican case does not lie in the automatic transposition of its content, but in the rationality that inspires it: public innovation is legitimate only when accompanied by reinforced obligations of documentation, oversight, evaluation, and accountability.

Regulation (EU) 2024/1689 classifies as high-risk systems, among others, AI systems used by competent authorities to perform individual assessments of the risk of crime and terrorism (Annex III). For these systems, the Regulation requires: the establishment of risk-management systems, data governance and management of training data, technical documentation, automatic event logging, transparency and information to users, human oversight, and levels of accuracy, robustness, and cybersecurity. This set of obligations constitutes the most advanced standard currently available for the legitimate use of AI in preventive public-security contexts.

The European Union Agency for Cybersecurity (ENISA) has published specific guidelines on the use of AI in cybersecurity, underscoring both the opportunities these technologies offer for incident detection and response and the risks inherent in AI itself, including specific vulnerabilities such as adversarial attacks, data poisoning, and model evasion. These technical risks have a direct legal dimension: a compromised or manipulated detection system may produce false negatives, allowing real threats to pass, or false positives, generating unjustified state interventions against innocent persons.

5.6. The Experience of CISA in the United States

The Cybersecurity and Infrastructure Security Agency (CISA) of the United States has developed an integral approach to the use of AI in the protection of critical infrastructures that offers relevant lessons from a comparative perspective. CISA has published guidelines on the responsible use of AI in the cybersecurity domain that establish principles of transparency, human oversight, privacy, and accountability.

CISA's experience illustrates the importance of distinguishing between defensive uses of AI (protection of one's own systems, intrusion detection, incident response) and offensive or surveillance uses (identification of actors, tracking of individuals, analysis of communications). While the former are less controversial from the standpoint of fundamental rights, the latter require additional legal guarantees and more robust oversight mechanisms.

5.7. United Nations Frameworks on Artificial Intelligence and Counter-Terrorism

The United Nations system has produced in recent years a body of work specifically oriented toward the relationship between artificial intelligence and terrorism prevention, which is directly applicable to the object of this study. The United Nations Office of Counter-Terrorism and UNICRI (2021) documented both the malicious use of these technologies by terrorist actors and their employment by state agencies for content detection and network analysis and concluded that technical capacity advances more rapidly than the frameworks of guarantees that should accompany it.

With greater legal density, the position paper of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism (OHCHR, 2025) examines the risks these systems project onto equality and non-discrimination, privacy, freedom of expression, due process, and the right to an effective remedy. The document underscores that international human-rights obligations bind States independently of the degree of automation of the decision, and that the delegation of functions to private providers does not displace that responsibility. It constitutes a relevant referent for Mexico, insofar as it originates from a mechanism of the universal system itself whose interpretive doctrine bears upon the fulfilment of the conventional obligations assumed by the State.

6. Case Studies: Judicial Control of Algorithmic Prevention

The normative frameworks reviewed enunciate principles, but it is in the judicial forum that one observes how those principles operate in the face of concrete systems. The three cases examined below share a common feature: in each of them a public authority deployed an algorithmic risk-assessment system and a court verified whether the normative authorization and the associated guarantees proved sufficient. Their interest for the object of this work does not reside in the identity of the regulated phenomenon, but in the structure of the judicial reasoning applied.

6.1. The SyRI Case (Netherlands, 2020)

The District Court of The Hague decided, on 5 February 2020, the challenge to the *Systeem Risico Indicatie* (SyRI), a state tool that cross-referenced databases from different administrations to generate risk reports on persons allegedly linked to social-security fraud. The court declared that the enabling legislation was incompatible with Article 8(2) of the European Convention on Human Rights, as it failed to satisfy the test of necessity in a democratic society and the required balance between the purpose pursued and the interference with private life (Rechtbank Den Haag, 2020).

Three elements of the reasoning are transferable. The first is that the legitimacy of the aim—the fight against fraud—did not suffice to validate the means. The second is that the opacity of the risk model, whose indicators were not disclosed, prevented both public scrutiny and individual defense,

which the court assessed as an insufficiency of guarantees and not as a discretionary technical decision. The third is that persons included in the risk registers were not notified, so that they could neither know nor challenge their classification. Transposed to the domain of cyberterrorism, where requirements of secrecy are even more intense, this last point anticipates the central problem: if the affectation remains invisible to the affected person, the right to a remedy is emptied of content.

6.2. *The Ligue des droits humains Case (Court of Justice of the European Union, 2022)*

This pronouncement is the most directly pertinent, as it examines an automated processing system designed precisely to prevent terrorist offences. In its judgment of 21 June 2022, the Grand Chamber of the Court of Justice analyzed the validity of Directive (EU) 2016/681 on the use of passenger name record data and admitted its compatibility with the Charter of Fundamental Rights only under a strict interpretation, recognizing that the regime entails serious interferences with the rights to private life and to the protection of personal data (Court of Justice of the European Union, 2022). Two precisions of the ruling have the value of a standard. The first is the exclusion of machine-learning systems capable of modifying the evaluation criteria without human intervention or review: the Court reasoned that such a capacity for self-adaptation prevents both the human reviewer and the affected person from knowing the reasons for a positive match, thereby emptying supervision and judicial control of content. The second is the requirement that the pre-established evaluation criteria be specific, reliable and non-discriminatory, and that every match be the object of individual review by non-automated means before producing any effect. For the object of this work, the teaching is twofold. On the one hand, it confirms that the counter-terrorism purpose, even when it constitutes an objective of general interest of the first order, does not function as an open authorization for mass processing or for the technical autonomy of the system. On the other hand, it offers an operational criterion infrequently found in soft-law instruments: the opacity derived from self-learning is not an inconvenience that must be mitigated, but a reason to exclude that architecture from the domain of state prevention.

6.3. *The Bridges Case (United Kingdom, 2020)*

The Court of Appeal of England and Wales declared unlawful the use of live automated facial recognition by South Wales Police on three convergent grounds: the absence of a sufficient legal basis delimiting who could be included in the watchlists and where the system could be deployed; the insufficiency of the data-protection impact assessment; and the failure to fulfil the duty to assess whether the software presented biases on grounds of sex or ethnic origin (Court of Appeal of England and Wales, 2020). The ruling is useful because it shows that the deficit of normative authorization is not located in the absence of a norm, but in its indeterminacy: codes of practice and internal procedures existed, yet none resolved the questions of the who and the where. That requirement of

precision coincides with what is sustained in section 4.1 and reinforces the conclusion that generic security clauses do not satisfy the principle of legality. The third ground adds a relevant procedural element: the evaluation of bias was not treated as a technical best practice, but as a legal obligation whose non-fulfilment vitiated the deployment.

6.4. Cross-Cutting Lessons for the Design of Algorithmic Prevention

Read together, the three cases reveal a stable argumentative pattern. The courts do not question the legitimacy of the preventive aim or the utility of the tool; what they examine is the density of the legal authorization, the effective possibility of reconstructing the functioning of the system in a concrete case, the existence of human review with a real capacity for contradiction, and the availability of a remedy for the classified person. When any of these elements is lacking, unlawfulness is declared independently of the technical performance of the system.

This pattern permits the formulation of a warning for the Mexican debate. Public discussion of artificial intelligence in security tends to be framed in terms of installed capacity, budget and technological acquisition. The cases examined show that the critical point is situated earlier: without a norm that determines purposes, subjects, data and controls, and without accessible avenues of challenge, the deployment remains exposed to a declaration of invalidity that may render the investment made useless. Rights-protective regulation does not operate here as a brake on innovation, but as a condition of its legal sustainability.

7. The Mexican Case and the Need for Rights-Protective Regulation

7.1. Existing Constitutional and Normative Framework

Mexico faces a complex reality in which technological vulnerabilities, growing dependence on digital services, risks to critical infrastructure and the presence of criminal dynamics with adaptive capacity converge. In that context, the idea of strengthening state capacities for detection and prevention through AI is understandable. Nevertheless, institutional necessity does not equate to automatic legal authorization.

The Mexican legal order possesses constitutional bases for affirming the state obligation of protection, yet still lacks a specific and sufficiently developed regulation on public algorithmic governance in highly sensitive preventive tasks. Essential questions persist: which authorities could use these systems, under what exact purposes, with what activation parameters, what type of data could be processed, what external controls should intervene, and what rights the persons eventually affected by AI-assisted decisions or classifications would have.

The Political Constitution of the United Mexican States establishes in its Article 6 the right of access to public information and the protection of personal data as fundamental rights. Article 16 of the Constitution enshrines the right to privacy and establishes that private communications are inviolable, except by resolution of a judicial authority and with the formalities provided by law. These constitutional provisions constitute the minimum floor of protection that any algorithmic system for the detection of cyberterrorism must respect. For its part, the Federal Penal Code does not contain an autonomous definition of cyberterrorism as a criminal offence, which obliges recourse to the general figure of terrorism (Article 139 FPC) and to computer-related offences (Article 211 bis FPC), with the difficulties this entails in terms of normative determination and foreseeability.

7.2. Competent Institutions and Current State of Capacities

In the cybersecurity domain, Mexico has various institutions with relevant competences, although their coordination is not always clear. The Secretariat of Security and Citizen Protection (SSPC) and the Federal Ministerial Police have competences in the investigation of cyber offences. The National Guard has a Scientific Division with capacities in cybersecurity and digital investigation. The Secretariat of National Defense (SEDENA) and the Secretariat of the Navy (SEMAR) have cyber-defense capacities within their respective spheres. The National Intelligence Centre (CNI) has strategic-intelligence functions that include cyberspace as a domain of attention.

This institutional distribution generates the risk of duplication of efforts, lack of coordination and gaps in the chain of command, especially in the response to incidents that simultaneously have dimensions of public security and national security. The National Cybersecurity Strategy published in 2017 established a general governance framework in the matter but was not accompanied by the specific legislation necessary for its effective implementation, nor did it expressly contemplate the use of AI systems in detection and prevention functions.

It is necessary to mention that the latter has been entirely neglected and needs updating, being the first and only public policy for the Mexican State in the matter (to date). The foregoing is mentioned solely in order to provide an understanding of the Mexican context.

7.3. The Normative Gap and Its Consequences

This absence of normative definition generates a double risk. On the one hand, it may provoke institutional paralysis or underutilization of potentially valuable tools. On the other, it may encourage improvised, technocratic or opaque implementations based more on technological availability than on constitutional legitimacy. Between both extremes, the only juridically sustainable way out is a rights-protective regulation that integrates from the outset requirements of legality, documentation, impact assessment, audit, human oversight and avenues of review.

In the Latin American context, Mexico is not the only country that faces this gap. A review of regional normative frameworks shows that, unlike the European Union or the United States, Latin America lacks specific regulations on the use of AI in public security and terrorism prevention. In Latin America there exist heterogeneous normative and public-policy developments in the fields of data protection and artificial intelligence; nevertheless, a consolidated regional regulation on the state use of AI in public security and terrorism prevention is still not observed.

This regional situation contrasts with the persistence of cyber incidents against governmental infrastructures and essential services, which evidences the need to strengthen institutional capacities and normative frameworks. Thus, the gap between the escalation of threats and the normative lag is one of the principal challenges faced by the legal systems of the region.

8. Minimum Criteria for a Legitimate Use of AI in This Matter

On the basis of the analysis developed, it is possible to sustain that the use of AI by public authorities for the detection and prevention of cyberterrorism would be juridically defensible only if certain minimum criteria are fulfilled in a cumulative manner.

8.1. *Legitimate, Specific and Delimited Purpose*

In the first place, there must exist a legitimate, specific and delimited purpose. The tool cannot be authorized for open formulas of security or general surveillance. It must be clearly defined whether its objective is to protect critical infrastructures, to strengthen incident response or to detect concrete threats against strategic systems. The principle of purpose limitation, enshrined both in the LGPDPPSO and in the international frameworks reviewed, requires that data may not be processed for purposes distinct from those originally declared. In the context of AI systems, this implies that the system must be designed for a specific function and must not be used to identify, monitor or classify persons for reasons other than the prevention of cyberterrorism.

8.2. *Sufficient Normative Basis*

In the second place, a sufficient normative basis is required. The regulation must specify competent authorities, categories of data, temporal limits, activation scenarios, documentation mechanisms and institutional controls. Without this, the use of AI becomes an administrative practice without clear legal contours. Following the standard of “quality of the law” developed by the ECtHR, the enabling norm must be accessible to the public, formulated with sufficient precision so that citizens can foresee, to the extent reasonably possible, the circumstances in which the authorities may have recourse to such measures, and contain adequate safeguards against abuse.

8.3. *Suitability, Necessity and Proportionality*

In the third place, it is indispensable to overcome an analysis of suitability, necessity and proportionality. The adoption of the system must be justified as adequate to the purpose pursued, necessary in the face of the absence of less intrusive means, and proportional in relation to the intensity of impact on fundamental rights. The proportionality test, widely developed by constitutional doctrine and applied by constitutional courts in the Ibero-American sphere, requires a reasoning that demonstrates:

1. that the system is apt to contribute to the preventive purpose pursued;
2. that there do not exist alternative means equally effective yet less restrictive of rights and;
3. that the benefits in terms of security are superior to the costs in terms of restriction of rights. This analysis must be carried out *ex ante*, before the deployment of the system, and reviewed periodically in order to evaluate whether the conditions that justified it remain valid.

8.4. *Meaningful Human Oversight*

In the fourth place, meaningful human oversight is indispensable. No system should by itself generate decisions that affect rights, activate investigations or produce relevant legal consequences without qualified and documented human review. UNESCO (2021) and the OECD (2023) coincide in underscoring that human oversight cannot be merely formal. It is not enough that a responsible official exists in the organizational chart; the oversight must be effective, which implies that the official must have the information, the training and the authority necessary to understand the recommendations of the system, to question them when necessary and to assume responsibility for the final decision. Regulation (EU) 2024/1689 operationalizes this requirement by establishing that high-risk AI systems must be designed in such a way that human operators can ignore, interrupt or revoke the outputs of the system in real time.

8.5. *Data Quality, Audit and Independent Validation*

In the fifth place, data quality, audit and independent validation must be guaranteed. This includes bias assessment, robustness testing, continuous review and protocols for suspension or correction in the face of structural errors. The NIST AI RMF (2023) establishes that the management of data quality is an essential component of the governance of AI systems. Training, validation and testing data must be representative of the domain of application, must be documented and must be subjected to periodic quality evaluations. In the context of cyberterrorism-detection systems, this implies that the data used must adequately reflect the current threat landscape and must not be biased toward particular types of actors, techniques or attack vectors.

8.6. Traceability and Explainability

In the sixth place, sufficient traceability and explainability must be assured. The authority must be able to reconstruct the functioning of the system in a concrete case and to justify how its output was incorporated into the institutional decision.

8.7. Regime of Challenge and Responsibility

In the seventh place, there must exist a clear regime of challenge and responsibility. The person affected by an AI-assisted action must have effective means of review and reparation. The administration cannot dilute its responsibility in the algorithm or in the technological provider. This principle of non-delegable responsibility of public authorities is recognized both by the OECD (2023) and by Regulation (EU) 2024/1689, which expressly establishes that the use of high-risk AI systems by public entities does not exempt those entities from their legal responsibility under applicable law. In the Mexican legal order, this principle finds support in the regime of state patrimonial liability provided for in Article 113 of the Political Constitution of the United Mexican States and in the Federal Law on the Patrimonial Liability of the State.

9. Toward Rights-Protective Regulation: Proposals for the Mexican Context

The analysis developed in the preceding sections makes it possible to formulate certain concrete proposals for advancing toward a rights-protective regulation of the use of AI by public authorities in the detection and prevention of cyberterrorism in Mexico. These proposals do not purport to be exhaustive; rather, they seek to orient the legislative and normative debate toward the minimum elements that a regulation of this type should incorporate.

A prior condition for any regulation of the use of AI in the prevention of cyberterrorism is the existence of a clear and precise legal definition of the conduct sought to be prevented. The absence of a specific criminal offence of cyberterrorism in the Federal Penal Code generates a normative indeterminacy that is incompatible with the principle of legality and hinders the delimitation of the preventive object of algorithmic systems. The Mexican legislature should consider the autonomous criminalization of cyberterrorism, following the standards of United Nations Security Council Resolution 1373 (United Nations Security Council, 2001).

The Mexican legal order requires a law or a specific normative framework that regulates the use of AI systems by public authorities in security and prevention functions. This norm should:

1. delimit with precision the competent authorities and their respective spheres of action;
2. establish a closed catalogue of purposes for which the use of AI is authorized;
3. define the types of data that may be processed by the systems, with express prohibitions regarding especially sensitive categories;
4. establish maximum periods of data retention and mechanisms of depuration;
5. create a regime of impact assessment and independent audit prior to the deployment of any system; and
6. establish mechanisms of effective judicial review of AI-assisted actions.

The governance of the use of AI in public security requires an independent supervisory body, endowed with the technical and legal competences necessary to evaluate compliance with normative requirements, conduct audits, receive complaints, and impose sanctions.

Given the cross-border nature of cyberterrorism, Mexico should advance in its accession to the Budapest Convention and in the negotiation of bilateral and multilateral cooperation agreements that permit information exchange, technical assistance, and the harmonization of standards for the use of AI in the prevention of cyberterrorism. Cooperation should not be limited to intelligence exchange but should extend to the joint development of technical standards, risk-assessment frameworks, and best practices in the design and deployment of algorithmic systems.

10. Conclusions

The use of artificial intelligence by public authorities to detect and prevent cyberterrorism cannot be presented as a legally neutral solution, nor as an instrument whose technical efficacy alone legitimates its deployment. It constitutes, on the contrary, one of the most delicate areas of technological public law, precisely because it places innovation at the very core of the State's preventive power.

The prevention of cyberterrorism requires real institutional capacities. The State needs tools to identify anomalies, protect critical infrastructures, and respond in a timely manner to complex threats. That necessity, however, does not authorize the substitution of legal rationality by a logic of opaque automation. Prevention remains legitimate only when it continues to be subject to legality, proportionality, oversight, and responsibility.

The review of international and comparative frameworks reveals a clear convergence: public uses of AI in sensitive contexts may be admitted only under reinforced conditions of governance and control. This convergence is visible in the UNESCO Recommendation (2021), the OECD

Recommendation (2023), the NIST AI RMF (2023), and Regulation (EU) 2024/1689, which share a common core of principles: legality, legitimate purpose, proportionality, human oversight, traceability, explainability, and responsibility.

That lesson is fully applicable to the Mexican case. If the country opts to advance in the state use of AI for the detection and prevention of cyberterrorism, it must do so by means of a specific, rights-protective, and technically serious regulation capable of preventing the promise of technological anticipation from becoming a sophisticated form of erosion of fundamental rights.

Comparative judicial experience reinforces this conclusion with a practical datum. In the three cases analyzed, the courts invalidated the deployment not because of deficiencies in the performance of the systems, but because of the insufficiency of the normative authorization, the opacity of the risk model, or the absence of avenues of challenge for the classified persons. The lesson for the Mexican legislature is that prior normative design does not constitute an obstacle to technological adoption, but the condition that makes it possible to sustain it over time.

The analysis of the Mexican case reveals three normative deficits that must be addressed as a matter of priority: the absence of a specific criminal offence of cyberterrorism, the lack of a specific regulation for the use of AI in public security, and the non-existence of an independent supervisory body with sufficient technical and legal capacities for the control of algorithmic systems. Overcoming these deficits is not merely a requirement of normative modernization, but a condition for the State to be able to exercise its preventive functions legitimately in the digital environment.

In the final analysis, the correct question is not whether the State should use artificial intelligence to confront complex digital risks, but under what constitutional conditions it may do so without disfiguring the democratic order it purports to protect. The sustainable answer is clear: only within an architecture of legality, audit, transparency, human oversight, and effective review.

11. References

- Boix Palop, A. (2020). Los algoritmos son reglamentos: La necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones. *Revista de Derecho Público: Teoría y Método*, 1, 223–270.
- Código Penal Federal. Cámara de Diputados.
- Constitución Política de los Estados Unidos Mexicanos. Cámara de Diputados.
- Conway, M. (2017). Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research. *Studies in Conflict & Terrorism*, 40(1), 77–98.

Cotino Hueso, L. (2023). Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos. In F. Balaguer Callejón & L. Cotino Hueso (Eds.), *Derecho público de la inteligencia artificial* (pp. 347–402). Fundación Manuel Giménez Abad.

Council of Europe. (2001). *Convention on Cybercrime* (Budapest Convention, ETS No. 185).

Court of Appeal of England and Wales. (2020). *R (on the application of Bridges) v. Chief Constable of South Wales Police*, [2020] EWCA Civ 1058.

Court of Justice of the European Union. (2022). *Ligue des droits humains ASBL v. Conseil des ministres* (Case C-817/19), ECLI:EU:C:2022:491.

Cybersecurity and Infrastructure Security Agency. (2023). *2023–2024 CISA roadmap for artificial intelligence*. U.S. Department of Homeland Security.

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.-M., Mingoia, J., & Shankar, R. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, 4(2), 1–21.

Denning, D. E. (2000). *Cyberterrorism* [Testimony before the Special Oversight Panel on Terrorism, Committee on Armed Services, U.S. House of Representatives].

European Court of Human Rights. (1978). *Klass and Others v. Germany* (Application no. 5029/71).

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.

European Union Agency for Cybersecurity. (2023). *Artificial intelligence and cybersecurity research*. ENISA.

Inter-American Commission on Human Rights, Special Rapporteur for Freedom of Expression. (2013). *Freedom of expression and the internet* (OEA/Ser.L/V/II. CIDH/RELE/INF. 11/13).

Ley Federal de Responsabilidad Patrimonial del Estado. (2004). *Diario Oficial de la Federación*.

Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados. (2017). *Diario Oficial de la Federación*.

Macdonald, S., Jarvis, L., & Lavis, S. (2022). Cyberterrorism today? Findings from a follow-on survey of researchers. *Studies in Conflict & Terrorism*, 45(8), 727–752.

Mayer Lux, L. (2018). Una definición de ciberterrorismo. *Revista Chilena de Derecho y Tecnología*, 7(2), 5–25.

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453.

Organization for Economic Co-operation and Development. (2023). *Recommendation of the Council on Artificial Intelligence*.

Office of the United Nations High Commissioner for Human Rights. (2025). *Protecting human rights while using artificial intelligence to counter terrorism* (PP/AI/CT). Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism.

Organization of American States. (1969). *American Convention on Human Rights*.

Rechtbank Den Haag. (2020). *NJCM et al. v. The State of the Netherlands* (SyRI), ECLI:NL:RBDHA:2020:1878.

Secretaría de la Presidencia. (2017). *Estrategia Nacional de Ciberseguridad*. Gobierno de México.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*.

United Nations. (2000). *United Nations Convention against Transnational Organized Crime*. United Nations Office on Drugs and Crime.

United Nations Office of Counter-Terrorism & United Nations Interregional Crime and Justice Research Institute. (2021). *Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes*.

United Nations Office on Drugs and Crime. (n.d.). *Module 14: Hacktivism, terrorism, espionage, disinformation campaigns and cyber warfare*. Education for Justice Initiative.

United Nations Security Council. (2001). *Resolution 1373 (2001)* (S/RES/1373).

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112.

Venkatachary, S. K., Prasad, J., Alagappan, A., Andrews, L. J. B., Raj, R. A., & Duraisamy, S. (2024). Cybersecurity and cyber-terrorism challenges to energy-related infrastructures: Cybersecurity frameworks and economics. Comprehensive review. *International Journal of Critical Infrastructure Protection*, 45, Article 100677.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4), 505–523.

Author's Note on the Use of Artificial Intelligence: *It is hereby clarified that the use of Generative Artificial Intelligence tools was strictly auxiliary, instrumental, and non-substantive in nature within the process of preparing the manuscript. Their use was limited to tasks involving the formal revision of the text, such as spelling correction, punctuation, accentuation, grammatical agreement, removal of repetition, improvement of syntactic fluency, stylistic consistency, and adaptation to an appropriate academic register. In this regard, the tool served as editorial support comparable to linguistic or stylistic proofreading, aimed at improving the clarity, readability, and formal accuracy of the manuscript, without replacing the author's intellectual work.*

More specifically, Generative Artificial Intelligence was not used to produce substantive content for the article, formulate the research problem, develop the theoretical framework, define the methodology, conduct the analysis, interpret findings, generate arguments, draft complete sections, introduce new ideas, produce conclusions, select bibliographic sources, or create references. All academic, conceptual, methodological, and argumentative decisions remained under the sole responsibility of the author, who reviewed and validated every formal modification made to the text and retained full control over such changes.

Furthermore, it should be emphasized that the use of the tool was not intended to fill gaps in content, develop sections that had not previously been drafted, or replace any part of the research, reading, critical analysis, or academic writing processes. The manuscript was conceived, structured, and developed by the author; technological assistance was employed solely at a subsequent stage of linguistic revision and formal standardization, with the purpose of strengthening the academic presentation of the text without altering its authorship, originality, scientific content, or intellectual responsibility.