

Uso de inteligencia artificial por autoridades públicas para detección y prevención del ciberterrorismo

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

Resumen

El uso de inteligencia artificial por autoridades para detectar y prevenir riesgos asociados al ciberterrorismo plantea una tensión central entre eficacia institucional y límites constitucionales. Estas herramientas permiten procesar grandes volúmenes de información, identificar anomalías y proteger infraestructuras críticas, pero su legitimidad no puede medirse únicamente por su utilidad operativa. En un Estado democrático de derecho, la prevención algorítmica exige base normativa suficiente, finalidad legítima, proporcionalidad, protección de datos personales, no discriminación, supervisión humana significativa, trazabilidad y control judicial. El artículo desarrolla un análisis jurídico-documental y comparado sobre el empleo de sistemas de inteligencia artificial en tareas de detección temprana, evaluación de riesgo y prevención del ciberterrorismo. Se examinan el Convenio de Budapest, las recomendaciones de la UNESCO y de la OCDE, el NIST AI Risk Management Framework y el Reglamento (UE) 2024/1689, junto con la experiencia de ENISA y CISA, los marcos de Naciones Unidas sobre inteligencia artificial y contraterroismo y tres decisiones judiciales que han sometido a control sistemas algorítmicos de riesgo. Se concluye que la inteligencia artificial sólo puede emplearse legítimamente cuando se integra en una arquitectura garantista de legalidad, auditoría, responsabilidad y revisión efectiva.

Palabras clave: inteligencia artificial, ciberterrorismo, autoridades públicas, derechos fundamentales, ciberseguridad, gobernanza algorítmica, infraestructuras críticas.

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, Doctor en Administración y Políticas Públicas, Posdoctorado en Derecho. Universidad Nacional Abierta y a Distancia de México (UnADM).

Para citar este artículo: Ramírez Castañeda, C. (2026), Uso de inteligencia artificial por autoridades públicas para detección y prevención del ciberterrorismo, Revista Internacional de Estudios sobre Terrorismo, nº18, pp.7-29.

Recibido
18/05/2026

Aceptado
24/08/2026

1. Introducción

La inteligencia artificial se ha convertido en una de las herramientas más relevantes de la transformación digital del sector público. Su incorporación en funciones estatales ha generado expectativas de eficiencia, anticipación y optimización en ámbitos tan diversos como la salud, la administración, la justicia, la seguridad y la gestión de infraestructuras críticas. En materia de ciberseguridad, esas expectativas se intensifican, ya que la IA ofrece capacidades de procesamiento masivo, detección de anomalías, clasificación de incidentes y apoyo a la toma de decisiones en escenarios caracterizados por velocidad, complejidad e incertidumbre.

Cuando dichas capacidades se orientan a la detección y prevención del ciberterrorismo, el debate adquiere una densidad jurídica singular. La cuestión ya no es únicamente si la tecnología puede mejorar la respuesta del Estado, sino si su utilización por autoridades públicas resulta compatible con el principio de legalidad, los derechos fundamentales y los límites materiales del poder preventivo. En ese contexto, la IA se sitúa en el cruce entre dos exigencias constitucionales: por un lado, la obligación del Estado de proteger a la población y asegurar la continuidad de servicios esenciales; por otro, la necesidad de que toda intervención pública permanezca sometida al control democrático y a las garantías propias del Estado de derecho.

La dificultad aumenta debido a la propia inestabilidad conceptual del ciberterrorismo. Se trata de una categoría frecuentemente utilizada de forma expansiva, mezclando ataques a infraestructura crítica, sabotaje, propaganda extremista en línea, campañas coordinadas de intimidación digital y otras manifestaciones de criminalidad tecnológica. Como señalan Macdonald et al. (2022), la comunidad académica no ha alcanzado un consenso definitivo sobre los elementos constitutivos del ciberterrorismo, lo que dificulta su tratamiento normativo y su distinción de figuras conexas como el hacktivismo, el espionaje digital o la ciberguerra. Esa amplitud conceptual es jurídicamente problemática, pues puede conducir tanto a la banalización del término como a respuestas estatales excesivas. Por ello, un análisis serio del uso de IA por autoridades debe comenzar por una delimitación cuidadosa del objeto preventivo.

El estudio se inserta en un debate doctrinal que ha ganado densidad en los últimos años. La literatura sobre gobernanza algorítmica ha mostrado que los sistemas automatizados de gestión de riesgos no son meros instrumentos auxiliares, sino dispositivos que reconfiguran la forma misma en que el poder público define, mide y anticipa el peligro (Yeung, 2018). En paralelo, el derecho público europeo e iberoamericano ha comenzado a discutir la naturaleza jurídica de los algoritmos empleados por la administración y el estándar de habilitación legal que su uso reclama (Boix Palop, 2020; Cotino Hueso, 2023). A ello se suma una producción institucional específica sobre inteligencia artificial y contrterrorismo, impulsada desde Naciones Unidas, que advierte sobre los riesgos de opacidad, sesgo y sobrevigilancia asociados a estas aplicaciones (UNOCT y UNICRI, 2021; ACNUDH, 2025). El presente trabajo dialoga con esa literatura y busca aportar una lectura situada en el ordenamiento mexicano.

El propósito del presente artículo es examinar bajo qué condiciones el uso de inteligencia artificial por autoridades para detectar y prevenir el ciberterrorismo puede considerarse jurídicamente legítimo. La tesis que se sostiene es que la validez de estas herramientas no deriva de su sola utilidad técnica. Su legitimidad depende de una arquitectura garantista basada en la legalidad, la finalidad legítima, la proporcionalidad, la protección de datos, la supervisión humana significativa, la trazabilidad, la auditoría independiente y la existencia de mecanismos efectivos de revisión y responsabilidad.

2. Metodología

La investigación es cualitativa, jurídico-documental y analítico-comparada. Se basa en la revisión de literatura académica especializada sobre ciberterrorismo, inteligencia artificial, función pública, gobernanza algorítmica, derechos fundamentales y gestión de riesgos tecnológicos. Asimismo, se examinan instrumentos jurídicos y marcos de referencia internacionales relevantes para valorar el uso de IA en contextos sensibles de prevención estatal.

En el plano nacional, se toman como referentes la Constitución Política de los Estados Unidos Mexicanos y el Código Penal Federal, con el fin de ubicar la discusión en el marco de los principios de legalidad, debido proceso, derechos humanos y persecución de amenazas graves. En el plano internacional y comparado, se revisan la Convención de las Naciones Unidas contra la Delincuencia Organizada Transnacional, el Convenio sobre la Ciberdelincuencia del Consejo de Europa, la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial, la Recomendación del Consejo de la OCDE sobre Inteligencia Artificial, el NIST AI Risk Management Framework y el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo.

La aproximación comparada no persigue trasladar mecánicamente soluciones extranjeras al contexto mexicano. Más bien, busca identificar principios funcionales y estándares normativos que permitan evaluar la juridicidad del uso estatal de sistemas algorítmicos en tareas preventivas. Para ello, el análisis se organiza a partir de cinco variables: finalidad legítima, base normativa suficiente, intensidad de afectación a derechos fundamentales, nivel de autonomía del sistema y existencia de controles ex ante, concomitantes y ex post.

Junto al análisis normativo, el trabajo incorpora tres decisiones judiciales que han sometido a control sistemas algorítmicos de evaluación de riesgo desplegados por autoridades públicas. Su selección responde a tres criterios: que el sistema examinado hubiera sido operado por un poder público con finalidades preventivas o de seguridad, que la resolución hubiera abordado de manera expresa la suficiencia de la habilitación legal o las garantías frente al tratamiento automatizado, y que el pronunciamiento proviniera de un órgano jurisdiccional cuya doctrina resulte trasladable, como estándar argumentativo, al examen de constitucionalidad propio del ordenamiento mexicano. Estas resoluciones no operan como precedentes vinculantes, sino como casos que permiten observar cómo las cinco variables señaladas se aplican en escenarios reales de litigio.

Desde el punto de vista epistemológico, la investigación adopta un enfoque jurídico crítico que reconoce que el derecho no es neutral frente a la tecnología. Como apuntan Danaher et al. (2017), las instituciones jurídicas tienden a rezagarse respecto del desarrollo tecnológico, y esa brecha puede generar vacíos normativos con consecuencias directas sobre la tutela de derechos. Por esta razón, el análisis no se limita a describir el estado actual de la regulación, sino que evalúa su suficiencia y formula propuestas de mejora orientadas a garantizar el ejercicio legítimo del poder preventivo del Estado.

3. Ciberterrorismo y prevención algorítmica: precisiones necesarias

3.1. Definición y delimitación del ciberterrorismo

La categoría de ciberterrorismo exige una delimitación cuidadosa. No todo incidente informático de alto impacto puede calificarse como tal, ni toda actividad criminal en el ciberespacio comporta la finalidad de intimidación colectiva o desestabilización institucional propia del terrorismo. El criterio decisivo no se encuentra en el medio digital empleado, sino en la finalidad perseguida y en la naturaleza de los bienes afectados.

La literatura especializada ha insistido en que el ciberterrorismo supone el uso de medios digitales con el objetivo de infundir miedo, coaccionar a autoridades o alterar funciones esenciales del Estado o de la vida social (Denning, 2000; Conway, 2017; Mayer Lux, 2018). Denning (2000) definió el ciberterrorismo como ataques ilegales y amenazas de ataque contra computadoras, redes y la información almacenada en ellas cuando se realizan para intimidar o coaccionar a un gobierno o a su población en la persecución de objetivos políticos o sociales. Esta definición, pionera en el campo, continúa siendo referencia obligada en la literatura especializada, si bien ha sido objeto de refinamientos posteriores.

Macdonald et al. (2022) realizaron una encuesta a investigadores especialistas en terrorismo y encontraron que existe un amplio consenso en torno a la necesidad de distinguir el ciberterrorismo del cibercrimen común y del hacktivismo sobre la base del elemento intencional: mientras el primero busca generar terror o desestabilización política, el segundo persigue ganancias económicas o la interrupción temporal de servicios, y el tercero tiene motivaciones protestatarias sin necesariamente producir daños de la magnitud propia del terrorismo. Esta distinción tiene implicaciones jurídicas de primer orden, pues determina el régimen aplicable, las garantías procesales exigibles y la proporcionalidad de las medidas preventivas.

Por su parte, Venkatachary et al. (2024) analizan los desafíos específicos que plantea el ciberterrorismo para las infraestructuras críticas del sector energético, identificando vectores de ataque diferenciados, marcos de ciberseguridad aplicables y dimensiones económicas del problema. Su análisis confirma que los ataques a infraestructuras esenciales pueden producir efectos en cadena

con consecuencias para la seguridad pública, la salud y la continuidad del Estado, lo que justifica una respuesta preventiva de alta intensidad, aunque sujeta a condiciones jurídicas estrictas.

La UNODC, en su módulo educativo sobre ciberdelincuencia (s. f.), distingue entre hacktivismo, terrorismo, espionaje, desinformación y ciberguerra como categorías analíticamente separadas, cada una con características propias respecto a actores, motivaciones, blancos y consecuencias jurídicas. Esta taxonomía es útil porque evidencia que la prevención algorítmica no puede operar con categorías indiferenciadas: el sistema debe estar diseñado para detectar conductas jurídicamente tipificadas, no para rastrear genéricamente “actividades sospechosas” sin contornos normativos precisos.

3.2. Prevención algorítmica: caracterización técnica y jurídica

En paralelo, la prevención algorítmica puede entenderse como el uso de sistemas computacionales capaces de identificar patrones, inferir riesgos, clasificar eventos o priorizar alertas para orientar la respuesta institucional. En contextos de ciberseguridad pública, esa prevención puede manifestarse en sistemas de detección de anomalías en redes, herramientas de correlación de incidentes, análisis automatizado de indicadores de compromiso o procesamiento de lenguaje en fuentes abiertas.

Desde una perspectiva técnica, los sistemas de IA aplicados a la ciberseguridad pueden clasificarse en diversas categorías según su función: sistemas de detección de intrusiones basados en aprendizaje automático (ML-IDS), plataformas de inteligencia de amenazas (Threat Intelligence Platforms, TIP), herramientas de análisis de comportamiento de entidades y usuarios (UEBA), sistemas de correlación de eventos de seguridad (SIEM potenciados con IA), y modelos de lenguaje natural para análisis de contenido en fuentes abiertas (OSINT). Cada tipo de sistema plantea condicionantes jurídicos distintos en términos de datos procesados, nivel de autonomía y potencial de afectación a derechos (European Union Agency for Cybersecurity, 2023; National Institute of Standards and Technology, 2023).

Conviene precisar, además, qué tipo de regulación algorítmica se pone en juego. Yeung (2018) propone distinguir entre sistemas reactivos, que responden a comportamientos ya verificados, y sistemas preventivos o anticipatorios, que actúan sobre inferencias acerca de conductas futuras, y advierte que estos últimos concentran los problemas de legitimidad más serios, pues intervienen antes de que exista un hecho jurídicamente relevante. La detección del ciberterrorismo se sitúa precisamente en ese segundo grupo. De ahí que el examen no pueda limitarse a la fiabilidad estadística del modelo: lo que está en discusión es la licitud de anticipar consecuencias jurídicas sobre la base de una probabilidad calculada.

Sin embargo, la complejidad jurídica aumenta cuando estos sistemas dejan de ser simples apoyos técnicos y comienzan a influir en decisiones institucionales sobre vigilancia, priorización de riesgos o apertura de líneas de investigación. La diferencia entre apoyo operativo y estructuración del juicio preventivo es central. En el primer caso, el sistema puede servir para fortalecer la resiliencia

de redes o infraestructuras. En el segundo, la IA empieza a incidir directamente sobre derechos, expectativas de privacidad, patrones de monitoreo o procesos administrativos y penales.

El Reglamento (UE) 2024/1689, conocido como Ley de Inteligencia Artificial, establece precisamente esta distinción como criterio de clasificación de riesgos. Los sistemas de IA utilizados por autoridades para la aplicación de la ley, la evaluación de riesgos o la identificación de sospechosos son considerados de alto riesgo y quedan sujetos a obligaciones reforzadas de transparencia, documentación, supervisión humana y evaluación de conformidad antes de su despliegue. Esta categorización normativa confirma que el derecho europeo ya reconoce que no todos los usos de IA en el ámbito público son equivalentes y que la intensidad regulatoria debe graduarse en función del impacto potencial sobre derechos fundamentales.

De ahí que la discusión jurídica no deba centrarse sólo en la sofisticación técnica del modelo, sino en el lugar institucional que ocupa dentro de la actuación estatal.

4. Riesgos jurídicos del uso de IA por autoridades públicas

4.1. Insuficiencia de base normativa

El primer riesgo jurídico es la insuficiencia de base normativa. Ninguna autoridad pública puede justificar el uso intensivo de IA en funciones preventivas sobre la base de cláusulas genéricas de seguridad o modernización administrativa. El principio de legalidad exige una habilitación clara respecto de finalidades, autoridades competentes, categorías de datos, límites temporales, supuestos de activación y mecanismos de supervisión. Sin una base normativa suficiente, la herramienta tecnológica termina sustituyendo a la norma y desdibujando la garantía de previsibilidad.

La doctrina administrativista ha ofrecido una vía para precisar esta exigencia. Boix Palop (2020) sostiene que los algoritmos empleados por la administración para adoptar decisiones cumplen una función materialmente equivalente a la de un reglamento, en tanto reglan y predeterminan la actuación del poder público, por lo que deberían quedar sujetos a las garantías de publicidad, participación y control ya previstas para las normas reglamentarias. Cotino Hueso (2023), al examinar los sistemas de identificación biométrica, subraya que la superposición entre la regulación de inteligencia artificial y la de protección de datos no dispensa de una habilitación legal específica y de calidad, y que los estándares exigidos por diversos tribunales constitucionales europeos rara vez se satisfacen en la práctica administrativa. Ambas aproximaciones convergen en un punto relevante para este trabajo: la reserva de ley no se agota en la existencia de una norma que mencione la tecnología, sino que reclama densidad normativa sobre su diseño, sus límites y sus controles.

El Tribunal Europeo de Derechos Humanos ha desarrollado en su jurisprudencia la doctrina de la “calidad de la ley” como condición para que las injerencias en derechos fundamentales sean compatibles con el Convenio Europeo. Según esta doctrina, la ley habilitante no sólo debe existir,

sino ser accesible, previsible y estar dotada de garantías adecuadas contra el abuso. Aunque esta jurisprudencia no es directamente vinculante para México, ofrece un estándar útil para evaluar la suficiencia de las habilitaciones normativas en materia de vigilancia y prevención algorítmica.

En el contexto latinoamericano, la Relatoría Especial para la Libertad de Expresión de la CIDH ha alertado sobre los riesgos de la vigilancia digital sin habilitación legal clara, señalando que la interceptación de comunicaciones y el procesamiento masivo de datos por parte del Estado requieren marcos normativos específicos que establezcan con precisión los supuestos, las autoridades competentes y los mecanismos de control. Aunque sus pronunciamientos se refieren principalmente a la vigilancia de comunicaciones, los principios que desarrollan son plenamente aplicables a los sistemas de detección algorítmica del ciberterrorismo.

4.2. Protección de datos personales

Un segundo riesgo radica en la protección de datos personales. La lógica preventiva tiende a favorecer la acumulación de información bajo la premisa de que más datos mejoran la capacidad de detección. Sin embargo, ese supuesto debe someterse a los principios de minimización, finalidad específica, proporcionalidad, exactitud y seguridad. La recolección masiva e indiscriminada no sólo es jurídicamente problemática, sino también técnicamente cuestionable, pues puede aumentar el ruido informacional y producir inferencias erróneas o excesivas.

En México, la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados establece principios rectores aplicables al tratamiento de datos por autoridades públicas, incluyendo licitud, finalidad, lealtad, proporcionalidad y responsabilidad. No obstante, la ley no contiene disposiciones específicas sobre el tratamiento de datos en sistemas algorítmicos de seguridad pública, lo que mantiene una zona de insuficiencia normativa que la regulación futura deberá atender.

La UNESCO (2021), en su Recomendación sobre la Ética de la IA, dedica especial atención a la protección de la privacidad y de los datos personales como condición necesaria para el desarrollo y despliegue ético de sistemas de IA. La Recomendación establece que los sistemas de IA no deben utilizarse para infringir el derecho a la privacidad, que incluye el control sobre los datos personales y la protección contra la vigilancia arbitraria o ilegal. Estos principios resultan directamente aplicables al diseño de sistemas de detección algorítmica del ciberterrorismo por parte de autoridades públicas.

4.3. Sesgo, discriminación y validez epistémica

En tercer lugar, debe considerarse el problema del sesgo y la discriminación. Los sistemas algorítmicos aprenden a partir de datos, decisiones de diseño y contextos institucionales. Si esos elementos están marcados por asimetrías, errores o lógicas de sobre-vigilancia, la herramienta puede reproducir o amplificar tales distorsiones. En materia de prevención del ciberterrorismo, ello

es especialmente grave, ya que una clasificación errónea puede activar vigilancia reforzada, afectar reputaciones o generar tratamientos desproporcionados antes de que exista un control judicial efectivo.

La literatura especializada ha documentado extensamente los riesgos de sesgo en sistemas de IA utilizados en contextos de seguridad. Estudios como los de Obermeyer et al. (2019) han demostrado que algoritmos utilizados en la toma de decisiones públicas pueden codificar y amplificar sesgos raciales, socioeconómicos o de género presentes en los datos de entrenamiento. En el ámbito específico de la seguridad, los sesgos pueden traducirse en sobre-vigilancia de determinadas comunidades, generando un impacto desproporcionado sobre personas y grupos que ya se encuentran en situación de vulnerabilidad. Esta dinámica es incompatible con el principio de igualdad y no discriminación, consagrado en los artículos 1º de la Constitución Política de los Estados Unidos Mexicanos y 24 de la Convención Americana sobre Derechos Humanos.

El NIST AI Risk Management Framework (2023) aborda el problema del sesgo como un riesgo sistémico que debe gestionarse en todas las etapas del ciclo de vida del sistema: diseño, desarrollo, validación, despliegue y monitoreo continuo. El marco propone mecanismos de evaluación de equidad, pruebas adversariales y mecanismos de retroalimentación para detectar y corregir sesgos, subrayando que la gestión del sesgo no es un evento único sino un proceso permanente. Esta perspectiva es especialmente relevante para los sistemas de detección del ciberterrorismo, donde la dinámica de las amenazas cambia constantemente y los datos de entrenamiento pueden volverse rápidamente obsoletos o sesgados.

4.4. Opacidad y exigencia de explicabilidad

Un cuarto riesgo es la opacidad. Cuando la autoridad no puede explicar suficientemente cómo operó el sistema, qué datos fueron relevantes o cómo se integró la recomendación algorítmica en la decisión final, la exigencia de motivación se debilita y el control judicial se vuelve precario. Por ello, la trazabilidad y la explicabilidad no son atributos deseables en abstracto, sino condiciones de juridicidad.

La OCDE (2023), en su Recomendación revisada sobre IA, establece que los sistemas de IA deben ser transparentes y que los operadores deben ser capaces de explicar sus decisiones y predicciones en la medida en que sea razonablemente posible. Asimismo, señala que los agentes de IA deberían poder proporcionar información suficiente para ayudar a las partes interesadas a comprender las capacidades y limitaciones de los sistemas y a responder a los resultados que producen. Estas exigencias de explicabilidad son especialmente relevantes en el contexto del uso estatal de IA, donde las decisiones tomadas con base en sistemas algorítmicos pueden afectar derechos fundamentales y deben poder ser objeto de revisión judicial efectiva.

El Reglamento (UE) 2024/1689 operacionaliza estas exigencias a través de obligaciones concretas para los sistemas de alto riesgo, incluyendo la llevanza de registros automáticos (logs),

la documentación técnica detallada, la provisión de instrucciones de uso, y la obligación de diseñar el sistema de manera que las salidas sean interpretables por las personas físicas responsables de su supervisión. Estas obligaciones no sólo tienen un valor instrumental para el control judicial, sino que contribuyen a construir la confianza institucional necesaria para el despliegue legítimo de la IA en contextos sensibles.

La doctrina ha señalado, sin embargo, los límites de este modelo. Veale y Zuiderveen Borgesius (2021) advierten que buena parte de las obligaciones de transparencia del régimen europeo descansa en mecanismos de autoevaluación de conformidad, de modo que el cumplimiento queda en manos del propio proveedor o del organismo implantador, con una supervisión externa limitada. Esta observación resulta especialmente pertinente en el ámbito de la seguridad, donde las exigencias de reserva pueden invocarse para restringir el acceso a la documentación técnica. Si la explicación se reduce a un expediente interno inaccesible para el afectado y para el juez, la garantía se vuelve nominal.

4.5. Responsabilidad pública y riesgo de tecnocracia

Finalmente, existe el riesgo de debilitamiento de la responsabilidad pública. La administración no puede escudarse en la complejidad técnica del sistema ni en el prestigio del proveedor para diluir su responsabilidad. Si la tecnología influye en una actuación estatal, la imputación jurídica sigue correspondiendo a la autoridad. La supervisión humana significativa, en este sentido, es también una exigencia de responsabilidad democrática.

El riesgo de tecnocracia se manifiesta cuando la complejidad técnica de los sistemas algorítmicos genera una asimetría informacional entre los operadores tecnológicos y los decisores políticos y judiciales. Esta asimetría puede producir lo que la doctrina denomina “deferencia algorítmica”: la tendencia de los funcionarios a aceptar sin cuestionamiento las recomendaciones de los sistemas automatizados, delegando efectivamente la decisión en el algoritmo. En contextos de ciberseguridad y prevención del terrorismo, donde la urgencia operativa es elevada, este riesgo se intensifica.

La Recomendación de la OCDE sobre IA (2023) aborda este problema exigiendo que los sistemas de IA sean diseñados de manera que permitan la supervisión, el reemplazo o la reversión por parte de operadores humanos. La UNESCO (2021) va más lejos al establecer que las decisiones que afecten significativamente a los derechos y libertades de las personas deben poder ser revisadas y, si fuera necesario, corregidas por seres humanos.

5. Referentes internacionales y comparados

5.1. *Convención de Naciones Unidas contra la Delincuencia Organizada Transnacional*

La Convención de las Naciones Unidas contra la Delincuencia Organizada Transnacional no regula directamente la inteligencia artificial ni el ciberterrorismo, pero aporta una arquitectura valiosa para la cooperación internacional frente a fenómenos complejos y transfronterizos. Su principal enseñanza es que la respuesta estatal frente a riesgos organizados no puede descansar únicamente en herramientas técnicas, sino en coordinación institucional, intercambio de información y fortalecimiento de capacidades. La Convención establece, en su artículo 29, la obligación de los Estados parte de adoptar, desarrollar o mejorar programas de capacitación para su personal, incluyendo la capacitación en técnicas de investigación compatibles con los derechos humanos. Esta disposición es relevante para el uso de IA en la prevención del ciberterrorismo, ya que implica que cualquier herramienta tecnológica debe ser acompañada por formación adecuada que garantice su uso dentro de los parámetros del Estado de derecho.

5.2. *Convenio de Budapest sobre Ciberdelincuencia*

El Convenio de Budapest ofrece, por su parte, un marco esencial en materia de cooperación penal y preservación de evidencia digital. Su relevancia en este estudio reside en recordar que la detección tecnológica no sustituye la juridicidad de la investigación. La IA puede contribuir a identificar señales tempranas o patrones de amenaza, pero cualquier actuación posterior debe quedar integrada en procedimientos compatibles con el debido proceso y con las reglas de obtención y preservación de evidencia. El Convenio de Budapest, en su Capítulo II, establece disposiciones específicas sobre medidas procesales que los Estados deben adoptar para la investigación de delitos informáticos, incluyendo la conservación expedita de datos almacenados, la divulgación expedita de datos de tráfico conservados y la interceptación de datos de contenido. Estas disposiciones deben entenderse como un piso mínimo de garantías que los sistemas de IA utilizados en la detección del ciberterrorismo no pueden ignorar.

5.3. *Marco ético de UNESCO y OCDE*

La Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial y la Recomendación de la OCDE sobre Inteligencia Artificial aportan principios directamente útiles para la evaluación del uso estatal de IA: respeto a los derechos humanos, privacidad, transparencia, explicabilidad, responsabilidad, seguridad, robustez y supervisión humana. Ambos instrumentos muestran que el sector público, especialmente en ámbitos sensibles como seguridad y justicia, debe operar con estándares más exigentes que los contextos puramente privados o experimentales.

La UNESCO (2021) introduce el concepto de “evaluación del impacto ético” como herramienta prospectiva para identificar los posibles efectos de los sistemas de IA sobre los derechos humanos antes de su despliegue. Esta herramienta es especialmente relevante para los sistemas de detección del ciberterrorismo, dado que la anticipación de los efectos sobre la privacidad, la igualdad y el debido proceso puede contribuir a diseñar sistemas más respetuosos con los derechos desde el origen, incorporando lo que la doctrina denomina “privacidad por diseño” (*privacy by design*).

La OCDE (2023), por su parte, ha actualizado su Recomendación sobre IA para reforzar los principios de rendición de cuentas y responsabilidad, señalando que los agentes de IA deberían ser responsables de la aplicación correcta de los principios relacionados con la IA. En el ámbito del sector público, esta responsabilidad adquiere una dimensión constitucional, puesto que el Estado no actúa como un agente privado más, sino como titular del poder público cuya actuación está sometida a los principios de legalidad, eficacia y servicio al interés general.

5.4. NIST AI Risk Management Framework

El NIST AI Risk Management Framework añade un enfoque operativo especialmente valioso: la gestión continua del riesgo. La gobernanza de la IA no debe concebirse como una autorización única, sino como un proceso permanente de identificación, evaluación, mitigación, monitoreo y revisión. En contextos preventivos, donde el riesgo y la incertidumbre cambian constantemente, esta lógica resulta indispensable.

El AI RMF 1.0 del NIST (2023) estructura la gestión del riesgo en cuatro funciones principales: Gobernar (Govern), Mapear (Map), Medir (Measure) y Gestionar (Manage). La función de Gobernar establece las políticas, procesos, procedimientos y prácticas para la gobernanza del riesgo de IA. La función de Mapear contextualiza los riesgos identificados en función del propósito y las partes interesadas. La función de Medir analiza, evalúa y monitorea los riesgos. Y la función de Gestionar ordena las respuestas de riesgo y aplica las medidas apropiadas. Este marco es aplicable directamente al diseño de sistemas de detección del ciberterrorismo por autoridades.

5.5. Reglamento (UE) 2024/1689 y la experiencia de ENISA

El Reglamento (UE) 2024/1689 confirma que, cuanto mayor es la sensibilidad del entorno de uso y mayor la posibilidad de afectación a derechos fundamentales, más densas deben ser las obligaciones regulatorias. Su valor comparado para el caso mexicano no está en la traslación automática de su contenido, sino en la racionalidad que lo inspira: la innovación pública sólo es legítima cuando va acompañada de obligaciones reforzadas de documentación, supervisión, evaluación y responsabilidad.

El Reglamento (UE) 2024/1689 clasifica como sistemas de alto riesgo, entre otros, los sistemas de IA utilizados por autoridades competentes para realizar evaluaciones individuales del riesgo de

delincuencia y terrorismo (Anexo III). Para estos sistemas, el Reglamento exige: establecimiento de sistemas de gestión de riesgos, gobernanza de datos y gestión de datos de entrenamiento, documentación técnica, registro de eventos automático, transparencia e información a los usuarios, supervisión humana, y niveles de exactitud, solidez y ciberseguridad. Este conjunto de obligaciones constituye el estándar más avanzado actualmente disponible para el uso legítimo de IA en contextos preventivos de seguridad pública.

La Agencia de la Unión Europea para la Ciberseguridad (ENISA) ha publicado directrices específicas sobre el uso de IA en ciberseguridad, subrayando tanto las oportunidades que ofrecen estas tecnologías para la detección y respuesta a incidentes como los riesgos propios de la IA, incluyendo vulnerabilidades específicas como los ataques adversariales, el envenenamiento de datos (*data poisoning*) y la evasión de modelos. Estos riesgos técnicos tienen una dimensión jurídica directa: un sistema de detección vulnerado o manipulado puede producir falsos negativos, dejando pasar amenazas reales, o falsos positivos, generando intervenciones estatales injustificadas sobre personas inocentes.

5.6. La experiencia de CISA en los Estados Unidos

La Agencia de Seguridad de Infraestructura y Ciberseguridad (CISA) de los Estados Unidos ha desarrollado un enfoque integral para el uso de IA en la protección de infraestructuras críticas que ofrece enseñanzas relevantes desde una perspectiva comparada. CISA ha publicado directrices sobre el uso responsable de IA en el ámbito de la ciberseguridad que establecen principios de transparencia, supervisión humana, privacidad y rendición de cuentas.

La experiencia de CISA ilustra la importancia de distinguir entre los usos defensivos de la IA (protección de sistemas propios, detección de intrusiones, respuesta a incidentes) y los usos ofensivos o de vigilancia (identificación de actores, seguimiento de individuos, análisis de comunicaciones). Mientras los primeros son menos controvertidos desde el punto de vista de los derechos fundamentales, los segundos requieren garantías jurídicas adicionales y mecanismos de supervisión más robustos.

5.7. Los marcos de Naciones Unidas sobre inteligencia artificial y contraterrorismo

El sistema de Naciones Unidas ha producido en los últimos años un cuerpo de trabajo específicamente orientado a la relación entre inteligencia artificial y prevención del terrorismo, que resulta directamente aplicable al objeto de este estudio. La Oficina de las Naciones Unidas contra el Terrorismo y el UNICRI (2021) documentaron tanto el uso malicioso de estas tecnologías por actores terroristas como su empleo por agencias estatales para la detección de contenidos y el análisis de redes, y concluyeron que la capacidad técnica avanza con mayor rapidez que los marcos de garantía que deberían acompañarla.

Con mayor densidad jurídica, el documento de posición del Relator Especial sobre la promoción y protección de los derechos humanos en la lucha contra el terrorismo (ACNUDH, 2025) examina los riesgos que estos sistemas proyectan sobre la igualdad y la no discriminación, la privacidad, la libertad de expresión, el debido proceso y el derecho a un recurso efectivo. El documento subraya que las obligaciones internacionales en materia de derechos humanos vinculan a los Estados con independencia del grado de automatización de la decisión, y que la delegación de funciones en proveedores privados no desplaza esa responsabilidad. Se trata de un referente relevante para México, en tanto proviene de un mecanismo del propio sistema universal cuya doctrina interpretativa incide sobre el cumplimiento de las obligaciones convencionales asumidas por el Estado.

6. Estudios de caso: el control jurisdiccional de la prevención algorítmica

Los marcos normativos revisados enuncian principios, pero es en sede jurisdiccional donde se observa cómo esos principios operan frente a sistemas concretos. Los tres casos que se examinan a continuación comparten un rasgo: en todos ellos una autoridad pública desplegó un sistema de evaluación algorítmica de riesgo y un tribunal verificó si la habilitación normativa y las garantías asociadas resultaban suficientes. Su interés para el objeto de este trabajo no reside en la identidad del fenómeno regulado, sino en la estructura del razonamiento judicial aplicado.

6.1. *El caso SyRI (Países Bajos, 2020)*

El Tribunal de Distrito de La Haya resolvió, el 5 de febrero de 2020, la impugnación del *Systeem Risico Indicatie* (SyRI), una herramienta estatal que cruzaba bases de datos de distintas administraciones para generar informes de riesgo sobre personas presuntamente vinculadas a fraude en materia de seguridad social. El tribunal declaró que la legislación habilitante era incompatible con el artículo 8.2 del Convenio Europeo de Derechos Humanos, al no superar el juicio de necesidad en una sociedad democrática ni el equilibrio exigible entre finalidad perseguida e injerencia en la vida privada (Rechtbank Den Haag, 2020).

Tres elementos del razonamiento resultan trasladables. El primero es que la legitimidad del fin, la lucha contra el fraude no bastó para validar el medio. El segundo es que la opacidad del modelo de riesgo, cuyos indicadores no fueron divulgados, impidió tanto el escrutinio público como la defensa individual, lo que el tribunal valoró como una insuficiencia de garantías y no como una decisión técnica discrecional. El tercero es que las personas incluidas en los registros de riesgo no eran notificadas, de modo que no podían conocer ni impugnar su clasificación. Trasladado al ámbito del ciberterrorismo, donde las exigencias de reserva son todavía más intensas, este último punto anticipa el problema central: si la afectación permanece invisible para el afectado, el derecho al recurso se vacía de contenido.

6.2. *El caso Ligue des droits humains (Tribunal de Justicia de la Unión Europea, 2022)*

Este pronunciamiento es el más directamente pertinente, pues examina un sistema de tratamiento automatizado diseñado precisamente para prevenir delitos de terrorismo. En su sentencia de 21 de junio de 2022, la Gran Sala del Tribunal de Justicia analizó la validez de la Directiva (UE) 2016/681, relativa al uso de los datos del registro de nombres de los pasajeros, y admitió su compatibilidad con la Carta de los Derechos Fundamentales únicamente bajo una interpretación estricta, reconociendo que el régimen entraña injerencias serias en los derechos a la vida privada y a la protección de datos (Tribunal de Justicia de la Unión Europea, 2022).

Dos precisiones del fallo tienen valor de estándar. La primera es la exclusión de los sistemas de aprendizaje automático capaces de modificar sin intervención ni revisión humana los criterios de evaluación: el tribunal razonó que tal capacidad de autoadaptación impide que el revisor humano y la persona afectada conozcan las razones de una coincidencia positiva, lo que priva de contenido a la supervisión y al control judicial. La segunda es la exigencia de que los criterios preestablecidos de evaluación sean específicos, fiables y no discriminatorios, y que toda coincidencia sea objeto de revisión individual por medios no automatizados antes de producir efecto alguno.

Para el objeto de este trabajo, la enseñanza es doble. Por una parte, confirma que la finalidad antiterrorista, aun siendo un objetivo de interés general de primer orden, no funciona como habilitación abierta para el tratamiento masivo ni para la autonomía técnica del sistema. Por otra, ofrece un criterio operativo poco frecuente en los instrumentos de *soft law*: la opacidad derivada del autoaprendizaje no es un inconveniente que deba mitigarse, sino una razón para excluir esa arquitectura del ámbito preventivo estatal.

6.3. *El caso Bridges (Reino Unido, 2020)*

El Tribunal de Apelación de Inglaterra y Gales declaró ilícito el uso de reconocimiento facial automatizado en vivo por la policía de Gales del Sur, por tres motivos convergentes: la ausencia de una base legal suficiente que delimitara a quién podía incluirse en las listas de vigilancia y dónde podía desplegarse el sistema, la insuficiencia de la evaluación de impacto en materia de protección de datos, y el incumplimiento del deber de valorar si el software presentaba sesgos por razón de sexo o de origen étnico (Court of Appeal of England and Wales, 2020).

El fallo es útil porque muestra que el déficit de habilitación normativa no se localiza en la ausencia de norma, sino en su indeterminación: existían códigos de práctica y procedimientos internos, pero ninguno resolvía las preguntas sobre el quién y el dónde. Esa exigencia de precisión coincide con lo sostenido en el apartado 4.1 y refuerza la conclusión de que las cláusulas genéricas de seguridad no satisfacen el principio de legalidad. El tercer motivo añade un elemento procedimental relevante: la evaluación del sesgo no fue tratada como una buena práctica técnica, sino como una obligación jurídica cuyo incumplimiento vició el despliegue.

6.4. Lecciones transversales para el diseño de la prevención algorítmica

Leídos en conjunto, los tres casos revelan un patrón argumentativo estable. Los tribunales no cuestionan la legitimidad del fin preventivo ni la utilidad de la herramienta; lo que examinan es la densidad de la habilitación legal, la posibilidad efectiva de reconstruir el funcionamiento del sistema en un caso concreto, la existencia de una revisión humana con capacidad real de contradicción y la disponibilidad de un recurso para la persona clasificada. Cuando alguno de estos elementos falta, la ilicitud se declara con independencia del rendimiento técnico del sistema.

Este patrón permite formular una advertencia para el debate mexicano. La discusión pública sobre inteligencia artificial en seguridad tiende a plantearse en términos de capacidad instalada, presupuesto y adquisición tecnológica. Los casos examinados muestran que el punto crítico se sitúa antes: sin una norma que determine finalidades, sujetos, datos y controles, y sin vías de impugnación accesibles, el despliegue queda expuesto a una declaración de invalidez que puede inutilizar la inversión realizada. La regulación garantista no opera aquí como un freno a la innovación, sino como condición de su sostenibilidad jurídica.

7. El caso mexicano y la necesidad de una regulación garantista

7.1. Marco constitucional y normativo existente

México enfrenta una realidad compleja en la que convergen vulnerabilidades tecnológicas, dependencia creciente de servicios digitales, riesgos a infraestructura crítica y presencia de dinámicas criminales con capacidad de adaptación. En ese contexto, la idea de fortalecer capacidades estatales de detección y prevención mediante IA resulta comprensible. No obstante, la necesidad institucional no equivale a habilitación jurídica automática.

El ordenamiento mexicano dispone de bases constitucionales para afirmar la obligación estatal de protección, pero carece todavía de una regulación específica y suficientemente desarrollada sobre gobernanza algorítmica pública en tareas preventivas de alta sensibilidad. Persisten preguntas esenciales: qué autoridades podrían utilizar estos sistemas, bajo qué finalidades exactas, con qué parámetros de activación, qué tipo de datos podrían procesarse, qué controles externos deberían intervenir y qué derechos tendrían las personas eventualmente afectadas por decisiones o clasificaciones asistidas por IA.

La Constitución Política de los Estados Unidos Mexicanos establece en su artículo 6° el derecho de acceso a la información pública y la protección de datos personales como derechos fundamentales. El artículo 16 constitucional consagra el derecho a la privacidad y establece que las comunicaciones privadas son inviolables, salvo mediante resolución de autoridad judicial y con las formalidades previstas en la ley. Estas disposiciones constitucionales constituyen el piso mínimo de protección que cualquier sistema de detección algorítmica del ciberterrorismo debe respetar. Por su parte, el Código

Penal Federal no contiene una definición autónoma del ciberterrorismo como tipo penal, lo que obliga a recurrir a la figura general del terrorismo (artículo 139 CPF) y a los delitos informáticos (artículo 211 bis CPF), con las dificultades que ello conlleva en términos de determinación y previsibilidad normativa.

7.2. Instituciones competentes y estado actual de las capacidades

En el ámbito de la ciberseguridad, México cuenta con diversas instituciones con competencias relevantes, aunque su coordinación no siempre es clara. La Secretaría de Seguridad y Protección Ciudadana (SSPC) y la Policía Federal Ministerial tienen competencias en materia de investigación de delitos cibernéticos. La Guardia Nacional cuenta con una División Científica con capacidades en ciberseguridad e investigación digital. La Secretaría de la Defensa Nacional (SEDENA) y la Secretaría de Marina (SEMAR) tienen capacidades de ciberdefensa dentro de sus respectivos ámbitos. El Centro Nacional de Inteligencia (CNI) tiene funciones de inteligencia estratégica que incluyen el ciberespacio como dominio de atención.

Esta distribución institucional genera el riesgo de duplicación de esfuerzos, falta de coordinación y vacíos en la cadena de mando, especialmente en la respuesta a incidentes que tienen simultáneamente dimensiones de seguridad pública y seguridad nacional. La Estrategia Nacional de Ciberseguridad publicada en 2017 estableció un marco general de gobernanza en la materia, pero no fue acompañada de la legislación específica necesaria para su implementación efectiva, ni contempló de manera expresa el uso de sistemas de IA en las funciones de detección y prevención.

Es menester mencionar que esta última mencionada se encuentra en el total olvido y se requiere una actualización necesaria, al ser la primera y única política pública para el Estado mexicano en la materia (hasta la fecha). Lo anterior es únicamente mencionado para tener un entendimiento del contexto de México.

7.3. La brecha normativa y sus consecuencias

Esta ausencia de definición normativa genera un doble riesgo. Por un lado, puede provocar parálisis institucional o subutilización de herramientas potencialmente valiosas. Por otro, puede alentar implementaciones improvisadas, tecnocráticas u opacas, basadas más en disponibilidad tecnológica que en legitimidad constitucional. Entre ambos extremos, la única salida jurídicamente sostenible es una regulación garantista que integre desde el origen exigencias de legalidad, documentación, evaluación de impacto, auditoría, supervisión humana y vías de revisión.

En el contexto latinoamericano, México no es el único país que enfrenta esta brecha. Una revisión de los marcos normativos regionales muestra que, a diferencia de la Unión Europea o Estados Unidos, América Latina carece de regulaciones específicas sobre el uso de IA en seguridad pública

y prevención del terrorismo. En América Latina existen desarrollos normativos y de política pública heterogéneos en materia de protección de datos e inteligencia artificial; sin embargo, todavía no se observa una regulación regional consolidada sobre el uso estatal de IA en seguridad pública y prevención del terrorismo.

Esta situación regional contrasta con la persistencia de incidentes cibernéticos contra infraestructuras gubernamentales y servicios esenciales, lo que evidencia la necesidad de fortalecer capacidades institucionales y marcos normativos, así es como se abre una brecha entre la escalada de las amenazas y el rezago normativo es uno de los principales desafíos que enfrentan los sistemas jurídicos de la región.

8. Criterios mínimos para un uso legítimo de IA en esta materia

A partir del análisis desarrollado, es posible sostener que el uso de IA por autoridades públicas para la detección y prevención del ciberterrorismo sólo sería jurídicamente defendible si se cumplen de manera acumulativa ciertos criterios mínimos.

8.1. *Finalidad legítima, específica y delimitada*

En primer lugar, debe existir una finalidad legítima, específica y delimitada. La herramienta no puede habilitarse para fórmulas abiertas de seguridad o vigilancia general. Debe quedar claramente definido si su objetivo es proteger infraestructuras críticas, fortalecer la respuesta a incidentes o detectar amenazas concretas sobre sistemas estratégicos. El principio de finalidad, consagrado tanto en la LGPDPPSO como en los marcos internacionales revisados, exige que los datos no puedan ser tratados para finalidades distintas a las originalmente declaradas. En el contexto de los sistemas de IA, esto implica que el sistema debe ser diseñado para una función específica y no debe utilizarse para identificar, monitorear o clasificar a personas por razones distintas a la prevención del ciberterrorismo.

8.2. *Base normativa suficiente*

En segundo lugar, se requiere una base normativa suficiente. La regulación debe precisar autoridades competentes, categorías de datos, límites temporales, supuestos de activación, mecanismos de documentación y controles institucionales. Sin ello, el uso de IA se convierte en una práctica administrativa sin contornos jurídicos claros. Siguiendo el estándar de “calidad de la ley” desarrollado por el TEDH, la norma habilitante debe ser accesible al público, estar formulada con suficiente precisión para que los ciudadanos puedan prever, en la medida de lo razonablemente posible, las circunstancias en que las autoridades pueden recurrir a tales medidas, y contener garantías adecuadas contra los abusos.

8.3. Idoneidad, necesidad y proporcionalidad

En tercer término, es imprescindible superar un análisis de idoneidad, necesidad y proporcionalidad. La adopción del sistema debe justificarse como adecuada al fin perseguido, necesaria frente a la ausencia de medios menos intrusivos y proporcional en relación con la intensidad de afectación a derechos fundamentales. El test de proporcionalidad, ampliamente desarrollado por la doctrina constitucional y aplicado por los tribunales constitucionales en el ámbito iberoamericano, exige una fundamentación que demuestre:

1. que el sistema es apto para contribuir a la finalidad preventiva perseguida;
2. que no existen medios alternativos igualmente eficaces, pero menos restrictivos de derechos; y
3. que los beneficios en términos de seguridad son superiores a los costos en términos de restricción de derechos. Este análisis debe realizarse ex ante, antes del despliegue del sistema, y revisarse periódicamente para evaluar si las condiciones que lo justificaron siguen siendo válidas.

8.4. Supervisión humana significativa

En cuarto lugar, la supervisión humana significativa es indispensable. Ningún sistema debería generar por sí solo decisiones que afecten derechos, activen investigaciones o produzcan consecuencias jurídicas relevantes sin revisión humana cualificada y documentada. La UNESCO (2021) y la OCDE (2023) coinciden en subrayar que la supervisión humana no puede ser meramente formal. No basta con que exista un funcionario responsable en el organigrama; la supervisión debe ser efectiva, lo que implica que el funcionario debe tener la información, la formación y la autoridad necesarias para entender las recomendaciones del sistema, cuestionarlas cuando sea necesario y asumir la responsabilidad de la decisión final. El Reglamento (UE) 2024/1689 operacionaliza esta exigencia estableciendo que los sistemas de IA de alto riesgo deben ser diseñados de manera que los operadores humanos puedan ignorar, interrumpir o revocar las salidas del sistema en tiempo real.

8.5. Calidad de datos, auditoría y validación independiente

En quinto lugar, deben garantizarse calidad de datos, auditoría y validación independiente. Ello incluye evaluación de sesgos, pruebas de robustez, revisión continua y protocolos de suspensión o corrección ante errores estructurales. El NIST AI RMF (2023) establece que la gestión de la calidad de los datos es un componente esencial de la gobernanza de los sistemas de IA. Los datos de entrenamiento, validación y prueba deben ser representativos del dominio de aplicación, deben estar documentados y deben ser sometidos a evaluaciones periódicas de calidad. En el contexto de los sistemas de detección del ciberterrorismo, esto implica que los datos utilizados deben reflejar adecuadamente el panorama de amenazas actual y no estar sesgados hacia determinados tipos de actores, técnicas o vectores de ataque.

8.6. Trazabilidad y explicabilidad

En sexto lugar, debe asegurarse trazabilidad y explicabilidad suficiente. La autoridad tiene que poder reconstruir el funcionamiento del sistema en un caso concreto y justificar cómo su salida fue incorporada a la decisión institucional.

8.7. Régimen de impugnación y responsabilidad

En séptimo lugar, debe existir un régimen claro de impugnación y responsabilidad. La persona afectada por una actuación asistida por IA debe contar con medios efectivos de revisión y reparación. La administración no puede diluir su responsabilidad en el algoritmo ni en el proveedor tecnológico. Este principio de responsabilidad indelegable de las autoridades públicas es reconocido tanto por la OCDE (2023) como por el Reglamento (UE) 2024/1689, que establece expresamente que el uso de sistemas de IA de alto riesgo por entidades públicas no exime a dichas entidades de su responsabilidad jurídica conforme al derecho aplicable. En el ordenamiento mexicano, este principio encuentra apoyo en el régimen de responsabilidad patrimonial del Estado previsto en el artículo 113 de la Constitución Política de los Estados Unidos Mexicanos y en la Ley Federal de Responsabilidad Patrimonial del Estado.

9. Hacia una regulación garantista: propuestas para el contexto mexicano

El análisis desarrollado en las secciones precedentes permite formular algunas propuestas concretas para avanzar hacia una regulación garantista del uso de IA por autoridades públicas en la detección y prevención del ciberterrorismo en México. Estas propuestas no pretenden ser exhaustivas, sino orientar el debate legislativo y normativo hacia los elementos mínimos que una regulación de este tipo debería incorporar.

Una condición previa a cualquier regulación del uso de IA en la prevención del ciberterrorismo es contar con una definición jurídica clara y precisa de la conducta que se pretende prevenir. La ausencia de un tipo penal específico de ciberterrorismo en el Código Penal Federal genera una indeterminación normativa que es incompatible con el principio de legalidad y dificulta la delimitación del objeto preventivo de los sistemas algorítmicos. El legislador mexicano debería considerar la tipificación autónoma del ciberterrorismo, siguiendo los estándares de la Resolución 1373 del Consejo de Seguridad de las Naciones Unidas (Consejo de Seguridad de las Naciones Unidas, 2001).

El ordenamiento mexicano requiere una ley o un marco normativo específico que regule el uso de sistemas de IA por autoridades públicas en funciones de seguridad y prevención. Esta norma debería:

1. delimitar con precisión las autoridades competentes y sus respectivos ámbitos de actuación;
2. establecer un catálogo cerrado de finalidades para las cuales el uso de IA está autorizado;
3. definir los tipos de datos que pueden ser procesados por los sistemas, con prohibiciones expresas respecto a categorías especialmente sensibles;
4. establecer períodos máximos de conservación de datos y mecanismos de depuración;
5. crear un régimen de evaluación de impacto y auditoría independiente previo al despliegue de cualquier sistema; y
6. establecer mecanismos de revisión judicial efectiva de las actuaciones asistidas por IA.

La gobernanza del uso de IA en seguridad pública requiere un órgano de supervisión independiente, dotado de las competencias técnicas y jurídicas necesarias para evaluar el cumplimiento de los requisitos normativos, realizar auditorías, recibir quejas y emitir sanciones.

Dada la naturaleza transfronteriza del ciberterrorismo, México debería avanzar en la adhesión al Convenio de Budapest y en la negociación de acuerdos de cooperación bilateral y multilateral que permitan el intercambio de información, la asistencia técnica y la armonización de estándares para el uso de IA en la prevención del ciberterrorismo. La cooperación no debe limitarse al intercambio de inteligencia, sino extenderse al desarrollo conjunto de estándares técnicos, marcos de evaluación de riesgos y mejores prácticas en el diseño y despliegue de sistemas algorítmicos.

10. Conclusiones

El uso de inteligencia artificial por autoridades públicas para detectar y prevenir el ciberterrorismo no puede ser presentado como una solución jurídicamente neutra ni como un instrumento cuya sola eficacia técnica legitima su despliegue. Se trata, por el contrario, de una de las áreas más delicadas del derecho público tecnológico, precisamente porque coloca a la innovación en el núcleo mismo del poder preventivo del Estado.

La prevención del ciberterrorismo exige capacidades institucionales reales. El Estado necesita herramientas para identificar anomalías, proteger infraestructuras críticas y responder con oportunidad frente a amenazas complejas. Sin embargo, esa necesidad no autoriza la sustitución de la racionalidad jurídica por una lógica de automatización opaca. La prevención sigue siendo legítima sólo cuando permanece sometida a legalidad, proporcionalidad, supervisión y responsabilidad.

La revisión de los marcos internacionales y comparados muestra una convergencia clara: los usos públicos de IA en contextos sensibles sólo pueden admitirse bajo condiciones reforzadas de gobernanza y control. Esta convergencia es visible en la Recomendación de la UNESCO (2021), la Recomendación de la OCDE (2023), el NIST AI RMF (2023) y el Reglamento (UE) 2024/1689, que comparten un núcleo común de principios: legalidad, finalidad legítima, proporcionalidad, supervisión humana, trazabilidad, explicabilidad y responsabilidad.

Esa lección resulta plenamente aplicable al caso mexicano. Si el país opta por avanzar en el uso estatal de IA para la detección y prevención del ciberterrorismo, deberá hacerlo mediante una regulación específica, garantista y técnicamente seria, capaz de evitar que la promesa de anticipación tecnológica se convierta en una forma sofisticada de erosión de derechos fundamentales.

La experiencia jurisdiccional comparada refuerza esta conclusión con un dato práctico. En los tres casos analizados, los tribunales invalidaron el despliegue no por deficiencias en el rendimiento de los sistemas, sino por insuficiencia de la habilitación normativa, opacidad del modelo de riesgo o ausencia de vías de impugnación para las personas clasificadas. La lección para el legislador mexicano es que el diseño normativo previo no constituye un obstáculo para la adopción tecnológica, sino la condición que permite sostenerla en el tiempo.

El análisis del caso mexicano revela tres déficits normativos que deben abordarse de manera prioritaria: la ausencia de un tipo penal específico de ciberterrorismo, la falta de una regulación específica para el uso de IA en seguridad pública, y la inexistencia de un órgano de supervisión independiente con capacidades técnicas y jurídicas suficientes para el control de los sistemas algorítmicos. Superar estos déficits no es sólo una exigencia de modernización normativa, sino una condición para que el Estado pueda ejercer legítimamente sus funciones preventivas en el entorno digital.

En definitiva, la pregunta correcta no es si el Estado debe usar inteligencia artificial para enfrentar riesgos digitales complejos, sino bajo qué condiciones constitucionales puede hacerlo sin desfigurar el orden democrático que pretende proteger. La respuesta sostenible es clara: sólo dentro de una arquitectura de legalidad, auditoría, transparencia, supervisión humana y revisión efectiva.

11. Referencias

Alto Comisionado de las Naciones Unidas para los Derechos Humanos. (2025). Protecting human rights while using artificial intelligence to counter terrorism (PP/AI/CT). Relator Especial sobre la promoción y protección de los derechos humanos y las libertades fundamentales en la lucha contra el terrorismo.

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.

Comisión Interamericana de Derechos Humanos, Relatoría Especial para la Libertad de Expresión. (2013). *Libertad de expresión e internet*. OEA/Ser.L/V/II. CIDH/RELE/INF. 11/13.

Constitución Política de los Estados Unidos Mexicanos. Cámara de Diputados.

Consejo de Seguridad de las Naciones Unidas. (2001). *Resolución 1373 (2001)*. S/RES/1373(2001).

- 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. En F. Balaguer Callejón y L. Cotino Hueso (Coords.), *Derecho público de la inteligencia artificial* (pp. 347-402). Fundación Manuel Giménez Abad.
- Court of Appeal of England and Wales. (2020). R (on the application of Bridges) v. The Chief Constable of South Wales Police, [2020] EWCA Civ 1058, sentencia de 11 de agosto de 2020.
- Council of Europe. (2001). *Convention on Cybercrime (Budapest Convention)*, ETS No. 185.
- 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, Judgment of 6 September 1978.
- 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.
- 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.
- Naciones Unidas. (2000). *Convención de las Naciones Unidas contra la Delincuencia Organizada Transnacional*. Oficina de las Naciones Unidas contra la Droga y el Delito.
- 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.

OCDE. (2023). *Recommendation of the Council on Artificial Intelligence*.

Oficina de las Naciones Unidas contra el Terrorismo y Instituto Interregional de las Naciones Unidas para Investigaciones sobre la Delincuencia y la Justicia. (2021). Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes. UNOCT-UNCCT y UNICRI.

Organización de los Estados Americanos. (1969). *Convención Americana sobre Derechos Humanos*.

Rechtbank Den Haag. (2020). NJCM y otros c. Estado de los Países Bajos (SyRI), ECLI:NL:RBDHA:2020:1878, sentencia de 5 de febrero de 2020.

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

Tribunal de Justicia de la Unión Europea. (2022). Ligue des droits humains ASBL c. Conseil des ministres, asunto C-817/19, ECLI:EU:C:2022:491, sentencia de la Gran Sala de 21 de junio de 2022.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.

UNODC. (s. f.). *Module 14: Hacktivism, terrorism, espionage, disinformation campaigns and cyber warfare*. Education for Justice Initiative.

Veale, M., y 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, 100677.

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

Nota de autor respecto al uso de Inteligencia Artificial: Se aclara que el uso de herramientas de Inteligencia Artificial Generativa tuvo un carácter estrictamente auxiliar, instrumental y no sustantivo dentro del proceso de preparación del manuscrito. Su empleo se limitó a tareas de revisión formal del texto, tales como corrección ortográfica, puntuación, acentuación, concordancia gramatical, depuración de repeticiones, mejora de fluidez sintáctica, uniformidad estilística y adecuación del registro académico. En este sentido, la herramienta funcionó como un apoyo editorial comparable a una revisión lingüística o de estilo, orientada a mejorar la claridad, legibilidad y corrección formal del manuscrito, sin sustituir el trabajo intelectual del autor. De manera específica, no se utilizó Inteligencia Artificial Generativa para producir contenido sustantivo del artículo, formular el problema de investigación, construir el enfoque teórico, definir la metodología, elaborar el análisis, interpretar hallazgos, generar argumentos, redactar apartados completos, incorporar ideas nuevas, producir conclusiones, seleccionar bibliografía ni crear referencias. Las decisiones académicas, conceptuales, metodológicas y argumentativas permanecieron bajo responsabilidad exclusiva del autor, quien revisó, validó y conservó el control sobre cada modificación formal realizada al texto. Asimismo, se precisa que el uso de la herramienta no tuvo por finalidad llenar vacíos de contenido, desarrollar secciones no elaboradas previamente ni reemplazar procesos propios de investigación, lectura, análisis crítico o escritura académica. El manuscrito fue concebido, estructurado y desarrollado por el autor; la asistencia tecnológica se empleó únicamente en una fase posterior de revisión lingüística y normalización formal, con el propósito de fortalecer la presentación académica del texto sin alterar su autoría, originalidad, contenido científico ni responsabilidad intelectual.