
¿Está México preparado institucionalmente para contener los riesgos en inteligencia artificial?¹

Alejandra Leyva
Independent

With
Apart Research

Abstract

Este proyecto propone el Mapa de Riesgos para la Soberanía de IA, aplicado al caso mexicano. El objetivo es evaluar si México cuenta con una base institucional suficiente para adoptar inteligencia artificial y si tiene capacidad de respuesta ante distintos escenarios de riesgo. Para ello se construyó una base de datos a partir de las acciones que el gobierno mexicano ha tomado en materia de IA. Estas acciones se clasificaron según la categoría y la vía de riesgo propuestas, se vincularon con su relación con la soberanía de IA y se evaluó su estado actual de implementación.

Los resultados muestran que la agenda pública se concentra principalmente en riesgos de corto plazo, en particular el uso malicioso de la IA y el manejo de datos públicos y privados. Los riesgos más estructurales (seguridad computacional, concentración de poder y pérdida de control) tienen una cobertura institucional considerablemente menor. El hallazgo principal es que México no sólo enfrenta un vacío regulatorio, sino un problema de dirección estratégica: la política pública sobre IA necesita pasar de una lógica reactiva a una aproximación preventiva, capaz de anticipar riesgos y fortalecer capacidades públicas.

¹ Research conducted at the [The Global South AI Safety Hackathon](#), June 2026

1. Introduction

El Sur Global avanza en inteligencia artificial desde una asimetría estructural profunda. **Produce talento técnico** que muchas veces termina migrando hacia centros de innovación en Estados Unidos, Europa o Asia; **genera enormes volúmenes de datos** sociales, culturales, financieros y biométricos que son capturados por plataformas globales; y **adopta sistemas de IA** desarrollados principalmente por actores externos que controlan modelos, infraestructura y criterios de diseño.

Esta asimetría afecta principalmente a aquellos países con capacidades institucionales limitadas y brechas regulatorias, de modo que dependen de las grandes empresas tecnológicas y su decisión sobre el desarrollo de IA. En este sentido, la concentración de poder en torno a sistemas avanzados de IA ha sido identificada como un riesgo relevante, especialmente cuando los beneficios, capacidades y recursos quedan distribuidos de forma desigual entre actores públicos, privados y sociales (Slattery et al., 2024).

Las preguntas que pretende resolver este análisis son: ¿cuál es la situación de México a nivel institucional respecto a la inteligencia artificial? y ¿qué necesita México para adoptar inteligencia artificial sin perder autonomía sobre sus datos, sus instituciones y sus decisiones públicas?

Estas preguntas parten de la idea de que la adopción de IA no puede evaluarse únicamente en términos de innovación, productividad o acceso tecnológico, sino también desde la capacidad del Estado para conservar control significativo sobre sus datos, infraestructura, proveedores tecnológicos, sistemas públicos, seguridad nacional y valores epistémicos. A esta capacidad se le llama aquí soberanía de IA, entendida como una forma de control legítimo y efectivo sobre tecnologías digitales críticas, sus condiciones de uso y sus efectos institucionales (Roberts, 2024; Pava et al., 2026).

Por otra parte, esto es relevante porque los sistemas de IA no sólo procesan información, sino que también ordenan, clasifican y priorizan formas de conocimiento. Desde enfoques decoloniales de la IA, se ha señalado que los sistemas tecnológicos pueden reproducir relaciones de poder, valores dominantes y formas específicas de entender el mundo, especialmente cuando son desarrollados desde contextos que no necesariamente representan a las sociedades donde después serán desplegados (Mohamed, Png, & Isaac, 2020).

Por ello, resulta importante el mantener esa independencia y en este análisis se pretende demostrar en qué punto se encuentra México, aunque consideramos que este framework podría aplicarse a otros país. De modo que, a la capacidad de un país para evitar que sus decisiones públicas, su información estratégica, sus sistemas institucionales y sus marcos de conocimiento queden excesivamente condicionados por sistemas externos se le llama soberanía IA (Belli & Gaspar, 2023) .

Se propone así un **Mapa de Riesgos para la Soberanía de IA**, aplicado al caso mexicano. Este marco se ha pensado de manera general a partir de otros marcos de clasificación de riesgos de IA, pero su objetivo principal es acotar escenarios de riesgo relevantes para países cuya principal vulnerabilidad no es necesariamente la que proviene de desarrollar modelos de frontera, sino recibir, adoptar y depender de tecnologías producidas. El mapa distingue entre categorías de riesgo y las vías de riesgo, o en otras palabras lo que te lleva a que el escenario final suceda. Esta distinción permite que, en una

segunda fase de estudio, a partir de las vías de riesgo se puedan desarrollar señales de alerta temprana para anticipar problemas, observar trayectorias de vulnerabilidad y construir rutas de mitigación más claras.

En las siguientes secciones se buscará, primero, explicar el Mapa de Riesgos para la Soberanía de IA por categoría y vía de riesgo. Posteriormente, se utilizará como base la información proveniente del Observatorio IA México, que contiene legislación, anuncios oficiales y casos judiciales relacionados con inteligencia artificial. A partir de esta información, se trabajó una base de datos en la que se analizaron dichas acciones de gobierno para distinguir **si ya se encontraban en funcionamiento, si permanecían como promesas o propuestas no implementadas, y qué tipo de riesgo cubrían dentro del mapa propuesto.** A cada registro se le añadió una variable de riesgo y también se evaluó si su contenido se relacionaba directamente con la soberanía de IA. Con base en esta clasificación, se presentará una distribución general de los datos para mostrar cómo se encuentra México de acuerdo con este mapa de riesgo y, finalmente, se ofrecerán recomendaciones de política pública aplicables a los vacíos identificados.

2. Related Work / Framework of analysis

Existen múltiples formas de clasificar los riesgos asociados con la inteligencia artificial, y cada mapeo parte de supuestos distintos sobre qué tipo de daño se busca observar, en qué escala ocurre, entre otros. Para este análisis, las clasificaciones existentes sirvieron como punto de partida, sin embargo, no todas las categorías resultan igualmente relevantes para economías del Sur Global. Por ello, el marco propuesto no replica directamente estas taxonomías, sino que las utiliza como referencia para construir una clasificación más ajustada a las vulnerabilidades que pueden observarse desde acciones de gobierno, legislación y política pública. (Slattery et al., 2024; AI Risk Explorer, n.d.; Hendrycks et al., 2025)

Para construir el marco propuesto, se pensó específicamente en el tipo de información disponible para este estudio, que en este caso son **acciones del gobierno relacionadas con inteligencia artificial.** A diferencia de otros ejercicios de clasificación que pueden construirse a partir de incidentes por ejemplo, este marco está pensado para identificar áreas de riesgo que puedan traducirse en política pública. Por esa razón, se excluyen aquellos riesgos que sólo afectarían directamente a países, empresas o laboratorios que están desarrollando modelos de frontera. A continuación, se muestran las categorías de riesgo consideradas en este análisis:

Figura 1. Mapa de Riesgos para la Soberanía de IA

Escenarios de riesgo	Vía de riesgo
Uso malicioso <i>Short term</i>	• fraude • usurpación de identidad y <i>deepfakes</i> • manipulación electoral e informativa • uso de IA por el crimen organizado • ataques a infraestructura crítica

Uso de datos públicos y privados <i>Corto a mediano plazo</i>	• uso de datos personales sin consentimiento • uso de datos gubernamentales e información estratégica por actores externos • explotación de propiedad intelectual y contenidos culturales • falta de trazabilidad sobre datos usados por modelos
Seguridad del sistema <i>Mediano plazo</i>	• introducir dependencia y opacidad en sistemas públicos • mantener el control operativo de la infraestructura crítica • uso indeseable de biométricos • regulación de proveedores
Concentración de poder <i>Mediano plazo</i>	• regular límites y acceso de frontier AI developers • regular el acceso a APIs • dependencia tecnológica de otros países
Afectación a la estructura económica <i>Mediano a largo plazo</i>	• dependencia productiva a empresas y gobiernos sin desarrollar capacidad local • establecer soluciones al desplazamiento producido por la automatización laboral • desigualdad en el acceso a la tecnología • baja captura de valor nacional concentrado fuera del país
Seguridad de cómputo <i>Mediano a largo plazo</i>	• inversión en recursos ya sea cómputo, talento, energía, infraestructura digital • acceso propio a hardware (GPUs) • regulación de nube extranjera, lugar y jurisdicción para el procesamiento de datos • estándares, condiciones, relaciones con proveedores y/o contratos cloud • centros de datos seguros, sostenibles y suficientes
Pérdida de control <i>Largo plazo</i>	• la capacidad de autonomía del Estado frente a la dependencia de proveedores • mecanismos de auditabilidad y rendición de cuentas ante riesgos • control para que las instituciones no estén condicionadas por sistemas externos • capacidad de reacción ante fallas, costos, o ataques que pudieran producir una afectación de servicios públicos

Con esta lógica, la tabla presenta las categorías de riesgo consideradas en el análisis, su horizonte temporal aproximado y las vías de riesgo que podrían llevar a cada escenario, con el fin de identificar dónde podrían surgir necesidades concretas de intervención pública.

3. Methods

Para este análisis se utilizó como punto de partida la información del Observatorio IA México, que sistematiza legislación, casos judiciales y anuncios oficiales relacionados con inteligencia artificial en México. A partir de esta información se construyó una base de datos propia para observar cómo se distribuyen actualmente las acciones institucionales vinculadas con IA y qué áreas de riesgo cuentan con menor cobertura pública.

El análisis incluyó únicamente legislación aprobada, legislación activa, casos judiciales y promesas públicas que ya operan o se encuentran en desarrollo. Para cada registro se revisó su contenido y se añadieron nuevas variables de clasificación: el escenario de riesgo al que pertenece, la vía de riesgo que busca atender, su horizonte temporal aproximado y su estado de implementación.

También se incorporó una variable de soberanía de IA. Para ello, se evaluó si cada registro contribuía directa o indirectamente a conservar control sobre datos, sistemas públicos, infraestructura, proveedores tecnológicos, información estratégica o decisiones institucionales. La codificación permite distinguir entre acciones generales de regulación de IA y medidas que podrían fortalecer la capacidad del Estado mexicano para adoptar inteligencia artificial.

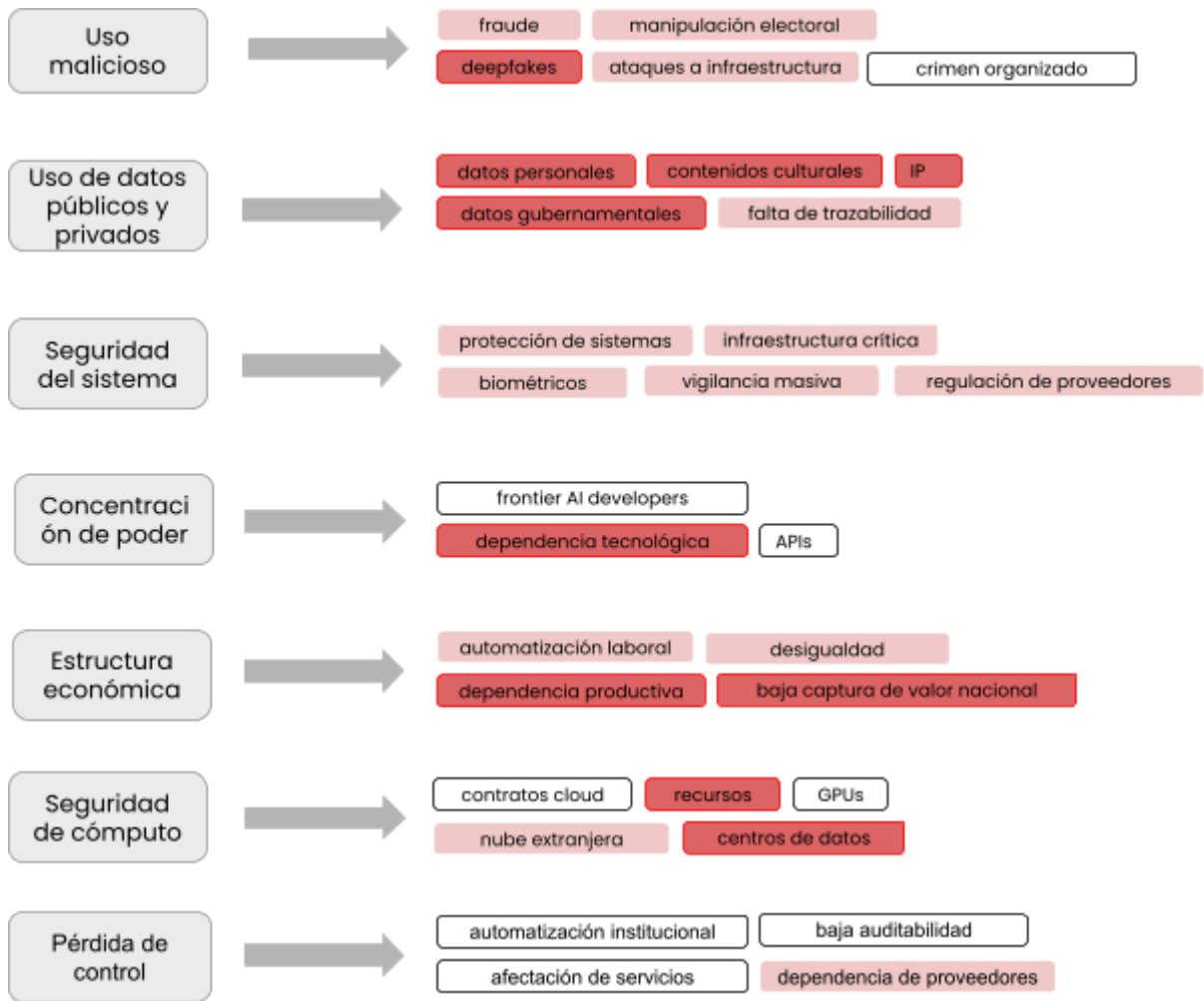
4. Results

El siguiente diagrama muestra una versión abreviada del Mapa de Riesgos para la Soberanía de IA. En oscuro se señalan aquellas vías de riesgo que ya aparecen cubiertas por legislación, casos judiciales o acciones gubernamentales en funcionamiento, mientras que en claro se muestran aquellas que han sido mencionadas en iniciativas, anuncios o promesas públicas, pero que todavía no se encuentran completamente implementadas. Las vías sin color muestran áreas donde no se identificó una acción institucional clara dentro de la base analizada (Figura 2).

El diagrama permite observar que las acciones actualmente en funcionamiento tienden a concentrarse en riesgos que ya han producido algún tipo de daño, queja o controversia pública, como deepfakes, uso de datos personales, contenidos culturales o propiedad intelectual. Esto sugiere que una parte importante de la respuesta institucional ha sido reactiva: el marco público parece avanzar con mayor claridad cuando el problema ya se ha vuelto visible y puede traducirse en sanciones, protección de derechos o mecanismos de denuncia. En contraste, otras áreas aparecen más como promesas o menciones generales, especialmente aquellas relacionadas con seguridad del sistema, regulación de proveedores, infraestructura crítica o vigilancia masiva. Esta diferencia es importante porque muestra que no todas las vías de riesgo tienen el mismo nivel de desarrollo institucional, incluso cuando ya han sido reconocidas como relevantes.

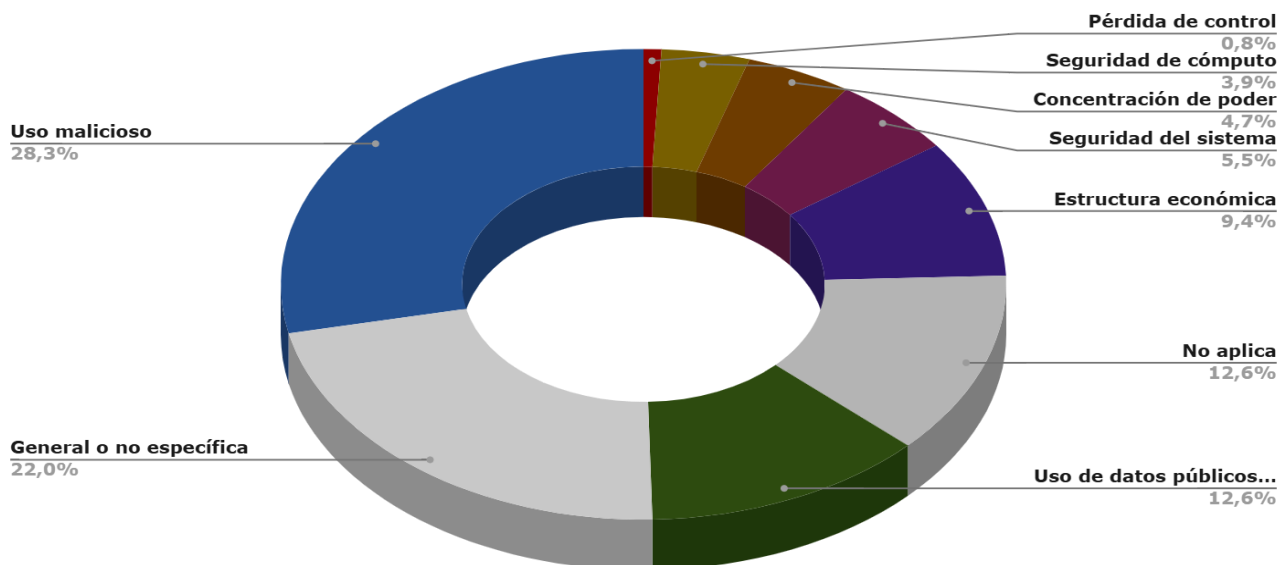
También llama la atención que los riesgos más estructurales, como la pérdida de control o la dependencia tecnológica de largo plazo, aparecen menos desarrollados en términos de acciones concretas. Esto no significa que sean riesgos imposibles de abordar, sino que todavía no parecen estar suficientemente traducidos en mecanismos institucionales claros. Precisamente por eso, el análisis sugiere que México podría anticiparse mejor si construye reglas, protocolos y capacidades públicas antes de que estos problemas se vuelvan más difíciles de corregir. Un buen diseño institucional podría permitir que algunos riesgos asociados con la soberanía de IA se atiendan desde etapas tempranas, especialmente si se conectan con obligaciones de auditabilidad, control público, rendición de cuentas y respuesta ante fallas.

Figura 2. Cobertura institucional de las vías de riesgo para la soberanía de IA en México



Gráfica 1. Distribución de acciones institucionales por categoría de riesgo

La gráfica de pastel muestra que casi una tercera parte de las acciones institucionales identificadas se concentra en uso malicioso de IA, lo cual parece responder a problemas inmediatos y visibles, como deepfakes, violencia sexual digital, contenido sexual que afecta a personas vulnerables y fraude. En segundo lugar aparece la categoría general o no específica, que agrupa leyes, iniciativas o anuncios que mencionan IA, seguridad o regulación de manera amplia, pero sin precisar qué riesgo buscan atender ni qué cambio institucional producirían. La categoría no aplica se utiliza para registros relacionados con temas administrativos o de uso interno de IA que no responden a un riesgo específico dentro del mapa, sino que se refieren más a procesos. También destaca el peso del uso de datos públicos y privados, especialmente por la presencia de temas vinculados con datos personales. En contraste, categorías más estructurales, como pérdida de control, seguridad de cómputo, concentración de poder o afectación a la estructura económica, aparecen con menor cobertura, lo que sugiere que la agenda pública todavía responde más a problemas de corto plazo que a riesgos preventivos o de largo plazo.

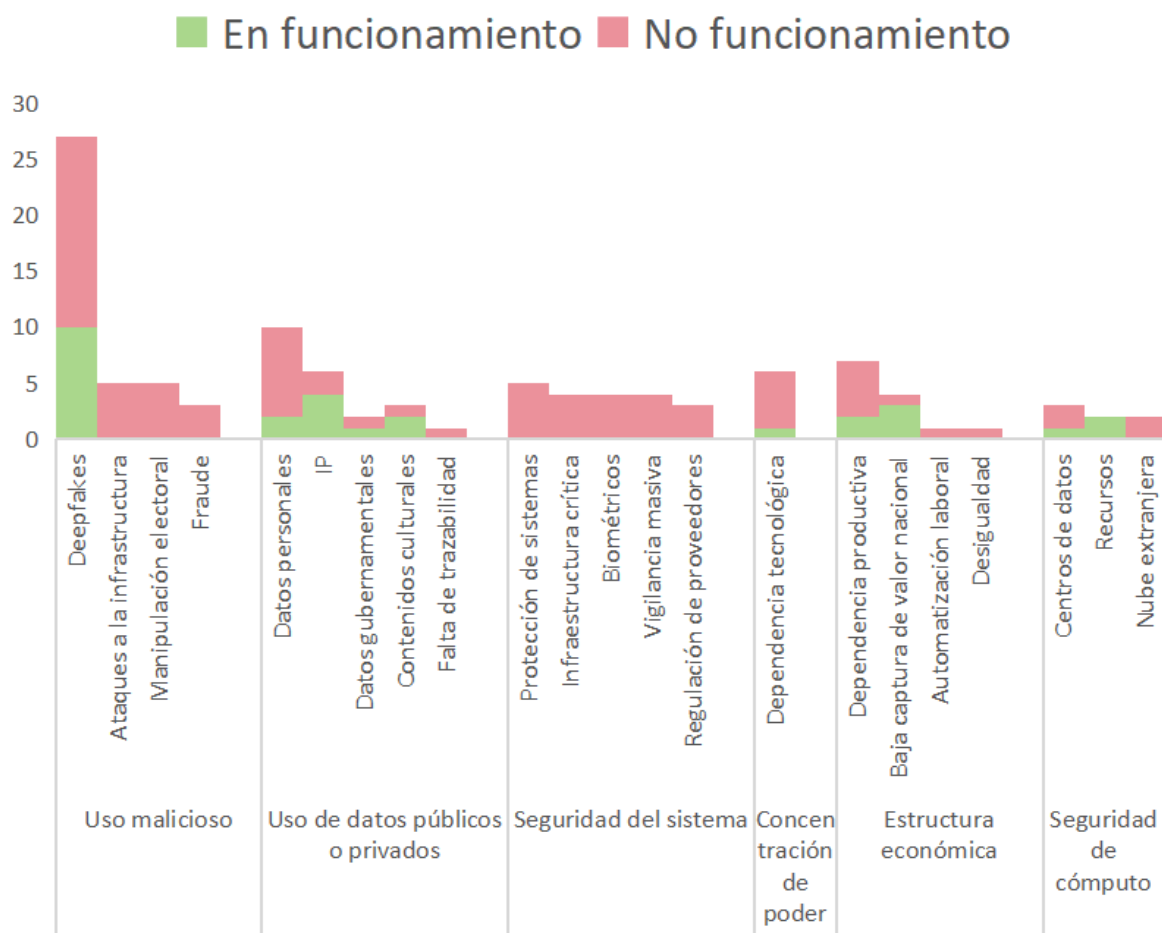


La siguiente gráfica muestra que una parte importante de las vías de riesgo identificadas todavía no se encuentra en funcionamiento. Aunque existen iniciativas o promesas en varias áreas, la cobertura operativa sigue siendo limitada. Esto es especialmente claro en seguridad del sistema sin acciones plenamente implementadas, si bien, puede leerse como una de las principales áreas de oportunidad, porque reúne riesgos relevantes para el funcionamiento del Estado y la protección de servicios públicos, pero todavía parece estar en una etapa principalmente declarativa.

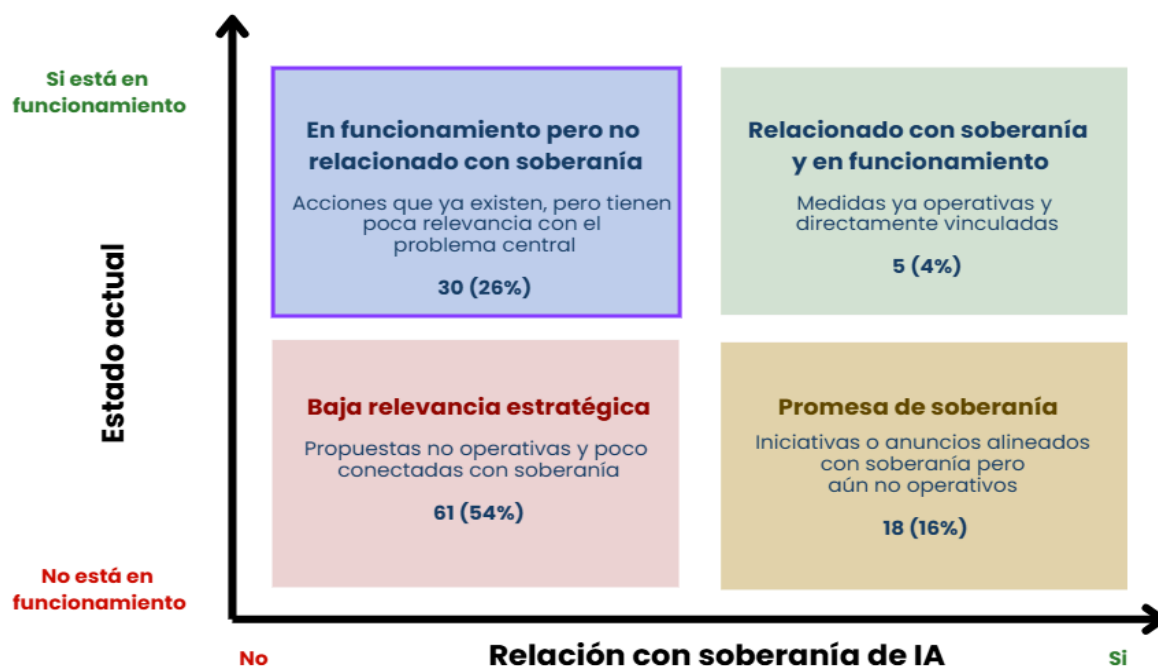
Dentro de uso malicioso, los deepfakes aparecen como la vía con mayor cobertura, lo que confirma que la respuesta institucional se ha concentrado en daños visibles y de corto plazo, especialmente vinculados con fraude, violencia sexual digital y manipulación de identidad. Así pues, la agenda pública ha avanzado más en riesgos inmediatos y denunciabiles, mientras que los riesgos estructurales de soberanía, que requieren planeación preventiva y capacidades estatales de largo plazo, permanecen menos desarrollados.

Gráfica 2. Estado de implementación por vía de riesgo

El siguiente hallazgo cruza dos dimensiones del análisis: si las acciones identificadas están actualmente en funcionamiento y si tienen una relación directa con la soberanía de IA. La mayor parte de los registros se concentra en el cuadrante de baja relevancia estratégica, que agrupa propuestas no operativas y poco conectadas con el problema central de soberanía. En contraste, solo una proporción reducida aparece como relacionada con soberanía y en funcionamiento, es decir, como medidas ya operativas y directamente vinculadas con la capacidad del Estado para conservar mayor control sobre datos, sistemas, infraestructura o decisiones institucionales.



Gráfica 3. Relación entre soberanía de IA y estado de implementación



Este resultado sugiere que el reto no es únicamente producir más iniciativas sobre IA, sino cambiar la dirección de la acción pública. El cuadrante verde tendría que aumentar con el tiempo, mientras que el peso de las acciones no operativas o poco conectadas con soberanía tendría que disminuir. En ese sentido, el mapa muestra una brecha importante entre la cantidad de actividad institucional relacionada con IA y la proporción de medidas que realmente fortalecen la soberanía de IA en México.

5. Discussion and Limitations

Este análisis debe leerse como una primera aproximación al nivel de preparación institucional de México frente a riesgos de soberanía de IA. Una limitación importante es que el estudio se concentró en registros donde la inteligencia artificial aparece de manera explícita, por lo que puede dejar fuera leyes, criterios judiciales o políticas públicas que, aunque no mencionen IA, podrían ser relevantes para contener estados adversos asociados con su uso.

Otra limitación es que el desarrollo de políticas públicas sobre inteligencia artificial en México todavía se encuentra en una etapa temprana. Por esa razón, el universo de observaciones plenamente aprobadas o en funcionamiento es reducido. Para ampliar el análisis, se incluyeron también iniciativas activas, anuncios oficiales y promesas públicas que ya operan o se encuentran en desarrollo. Esto permite observar hacia dónde parece moverse la agenda pública, pero también implica que algunos registros analizados podrían no llegar a implementarse. En ese sentido, el análisis toma en cuenta propuestas que todavía no producen efectos institucionales concretos.

Finalmente, la relación con soberanía de IA se codificó de forma binaria, es decir, como relacionada o no relacionada con soberanía. Esta decisión permitió hacer una primera clasificación clara, pero simplifica una realidad más gradual. Algunas medidas pueden estar directamente vinculadas con soberanía, mientras que otras pueden tener una relación más indirecta.

Futuros estudios podrían ampliar la capa metodológica de este análisis, tanto en la forma de medir la variable de soberanía como en la muestra utilizada. Una primera extensión importante sería incorporar una base de incidentes, no necesariamente limitada a México, sino también a países con condiciones institucionales y tecnológicas similares. Esto permitiría identificar casos concretos en los que una vía de riesgo ya se está materializando, lo que ayudaría a conectar el mapa de riesgos con señales más claras de alerta temprana.

Una segunda línea de trabajo sería traducir el marco a recomendaciones de política pública más puntuales para legisladores y tomadores de decisión que todavía enfrentan una curva de aprendizaje alta sobre estos temas. Esto podría facilitar mayor comunicación entre investigación técnica, análisis de riesgos y diseño institucional, así como abrir espacios de cabildeo público mejor informados.

Finalmente, también sería importante comunicar esta clasificación al público de manera accesible. Explicar qué incidentes pueden indicar la aparición de una vía de riesgo, por qué esos riesgos importan y por qué el tema de la soberanía es importante.

6. Conclusion

Se desarrolló un Mapa de Riesgos para la Soberanía de IA como una herramienta original para analizar la preparación institucional de un país que no se encuentra en el centro del desarrollo tecnológico global, pero que sí puede verse profundamente afectado por sus consecuencias. Al poner la soberanía de IA como eje del análisis, el estudio busca mostrar que la adopción de inteligencia artificial no solo implica acceso a nuevas herramientas, sino también decisiones que nos afectan y que pueden quedar relegadas únicamente a actores externos. Hacer visible esta dimensión resulta importante para entender qué se está entregando, bajo qué condiciones y qué podría implicar en escenarios de riesgo para la vida pública y cotidiana.

Los resultados muestran que la agenda pública mexicana ha avanzado principalmente en temas de uso malicioso, uso de datos públicos y privados y, en menor medida, seguridad del sistema, aunque muchas de estas acciones todavía permanecen como promesas o iniciativas no implementadas. Por otro lado, hay una tendencia punitiva y reactiva en la legislación, donde la respuesta institucional parece surgir después de un suceso. Por ello, una conclusión central es que la regulación de IA tendría que moverse hacia una lógica más preventiva, con una visión más amplia del riesgo y con capacidades públicas capaces de anticipar problemas antes de que se vuelvan más difíciles de corregir.

Code and Data

Base de datos:
https://docs.google.com/spreadsheets/d/18UHdCzaj4im6zUJ3rHoEJD8Pf--_N4TSyXT_IL7iHao/edit?usp=sharing

References

AI Risk Explorer. (2026). AI Risk Explorer. <https://www.airiskexplorer.com/>

AI Risk Observatory. (2026). AI Risk Observatory: Data dashboard. <https://www.riskobservatory.ai/data>

Belli, L., & Gaspar, W. B. (2023). AI transparency, AI accountability, and AI sovereignty: An overview. In L. Belli & W. B. Gaspar (Eds.), *The quest for AI sovereignty, transparency and accountability: Official outcome of the UN IGF Data and Artificial Intelligence Governance Coalition* (p. 23). FGV Direito Rio. https://aigovernance.fgv.br/sites/aigovernance.fgv.br/files/2024-01/the_quest_for_ai_sovereignty_transparency_and_accountability.pdf

Gladstone AI. (2026). Superintelligence strategy: Deterrence with Mutual Assured AI Malfunction. <https://www.nationalsecurity.ai/chapter/deterrence-with-mutual-assured-ai-malfunction-maim>

Hendrycks, D., Schmidt, E., & Wang, A. (2025). *Superintelligence strategy. Expert version*. arXiv.

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659–684. <https://doi.org/10.1007/s13347-020-00405-8>

Observatorio IA México. (2026). IA en el Estado mexicano. Recuperado el 21 de junio de 2026, de <https://www.observatorio-ia-mexico.com/>

Pava, J. N., Meinhardt, C., Cryst, E., & Landay, J. A. (2026, February 17). AI sovereignty's definitional dilemma. Stanford Institute for Human-Centered Artificial Intelligence. <https://hai.stanford.edu/news/ai-sovereignty-definitional-dilemma>

Roberts, H. (2024). Digital sovereignty and artificial intelligence: A normative approach. *Ethics and Information Technology*, 26, Article 70. <https://doi.org/10.1007/s10676-024-09810-5>

Slattery, P., Saeri, A. K., Grundy, E. A. C., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., & Thompson, N. (2024). The AI Risk Repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence. *arXiv*. <https://arxiv.org/abs/2408.12622>

LLM Usage Statement

Main idea, database creation, conclusions, text, graphs are owned. Claude was used to find errors in database, connectors, synonyms or correct grammar and syntax mistakes.