COMENTARIOS AL CUESTIONARIO DEL IMPACTO DE LA IA EN LOS DDHH EN LAS AMÉRICAS

R3D, Red en Defensa de los Derechos Digitales, es una organización mexicana no gubernamental dedicada a la defensa y promoción de los derechos humanos en el entorno digital. El presente escrito es nuestra contribución al cuestionario presentado por la H. Relatoría Especial para la Libertad de Expresión. En específico, respondemos a las preguntas primera, segunda, cuarta, sexta y séptima del cuestionario. Además de proporcionar evidencia sobre el uso y regulación de sistemas de inteligencia artificial, planteamos problemáticas específicas respecto al uso de estas herramientas parte del estado mexicano y proponemos marcos de análisis de derechos humanos que permitan problematizar el impacto de la IA en la libertad de expresión y su relación con otros derechos humanos.

I. Uso de inteligencia artificial por parte del Estado mexicano

En México, existe evidencia de que al menos 14 dependencias del gobierno utilizan sistemas de IA o están en proceso de desarrollarla internamente para incorporar como parte de sus funciones y servicios. Conforme a la información disponible, los ámbitos en que se usa son: vigilancia y seguridad e inteligencia,[1] servicios de atención ciudadana y trámites, asuntos fiscales y detección de irregularidades. En estos casos, la IA se utiliza sin la posibilidad de conocer sus finalidades y alcance, de qué manera es implementada y los límites o salvaguardas para prevenir abusos y violaciones a los derechos humanos, dado que dicha información se ha clasificado como reservada por las autoridades que la utilizan para fines públicos, por motivos de seguridad nacional.[2]

---

1. Entre otros, las autoridades utilizan sistemas predictivos para vigilancia en el espacio público (reconociimento facial). sistemas para detectar irregularidades o patrones de riesgo (alertas y perfiles fiscales) y uso de sistemas de IA para analizar bases de datos gigantescas para prevención de delitos e inteligencia (de acuerdo a lo habilitado por el paquete legislativo de vigilancia, como explicamos más adelante). Sin embargo, el principal problema es que la ausencia de obligaciones de transparencia, debido proceso y rendición de cuentas (establecida por la ausencia de legislación y por la legislación que establece la opacidad como regla general) hacen que no sea posible evaluar comprehensivamente los impactos que estas tecnologías tienen. Desarrollamos este punto en seguida.

2. En este sentido, “La Guardia Nacional (GN) acepta la integración de IA, pero no precisa su propósito. Aún más reservadas son la Unidad de Inteligencia Financiera (UIF) y la Fiscalía General de la República (FGR), que clasificaron la información. La UIF argumenta que revelar su uso representaría una amenaza a la seguridad nacional y pondría en riesgo las actividades de prevención de delitos y financiamiento al terrorismo. La FGR señala que al hacer pública su información podría exponerse a ataques para manipular los modelos de IA.”, ver en: Estrada, Andrés.”El gobierno mexicano ya usa la IA, pero no hay normas ni estrategia”, WIRED, 3 de marzo de 2026. Disponible en: https://es.wired.com/articulos/el-gobierno-mexicano-ya-usa-la-ia-pero-no-hay-normas-ni-estrategia

México no cuenta con una regulación general ni particular sobre el desarrollo, implementación y evaluación de las tecnologías de IA en México. Si bien existen legislaciones que facultan a autoridades específicas para utilizarla, existe un vacío legal de protección que garantice el cumplimiento de obligaciones mínimas de transparencia, debido proceso y respeto a los derechos humanos.

Diversos marcos de protección de derechos se ven afectados por la falta de transparencia en el ámbito público; por ejemplo las leyes de datos personales en manos de sujetos obligados y la correlativa a las obligaciones de los entes privados, así como los derechos al debido proceso y a la no discriminación. La falta de transparencia significativa (no sólo en términos contractuales y operativos sino también de carácter algorítmico, para saber qué tipo de datos, con qué procesos, para qué fines y con qué salvaguardas estos sistemas utilizan la información protegida por las leyes de datos personales) hace que, al menos en los ámbitos relacionados con la seguridad, la vigilancia estatal, el ámbito fiscal y financiero, sea imposible saber si las autoridades cumplen o no con las obligaciones jurídicas que existen actualmente.

a. Sistemas predictivos en materia de seguridad pública

En México, diversas autoridades de seguridad y defensa integran sistemas predictivos basados en inteligencia artificial para fines de prevención, inteligencia y operaciones militares y de ciberseguridad. R3D identifica como preocupante el siguiente marco normativo que habilita el uso de dichas herramientas:

- La Ley del Sistema Nacional de Investigación e Inteligencia en Materia de Seguridad Pública (en adelante, Ley de Inteligencia) crea una "Plataforma Central de Inteligencia" que se interconecta con distintas bases de datos públicas y privadas, generando un nodo único de información. La recolección masiva y centralizada de datos personales es preocupante en función de las facultades legales que la ley otorga a las autoridades de seguridad pública. El artículo 32 establece que las autoridades pueden utilizar herramientas de inteligencia artificial para procesar y analizar información recabada o utilizada por la plataforma con el fin de "elaborar productos de inteligencia" (énfasis propio). La Plataforma deberá entrar en funcionamiento a inicios de abril de 2026 conforme a los artículos transitorios de la Ley de Inteligencia.
- La Ley Orgánica de la Armada de México,[3] reformada en noviembre de 2025, establece en su artículo 2, numeral XIX, que la Armada tiene la facultad de "participar en

3. Decreto por el que se expide la Ley Orgánica de la Armada de México. Diario Oficial de la Federación. 7 de noviembre de 2025. Disponible en: https://www.diputados.gob.mx/LeyesBiblio/iniclave/66/CD-LXVI-II-1P-056/03_dof_056_07nov25.pdf

actividades de ciberdefensa y ciberseguridad para la conducción de las operaciones militares que se realizan en el ciberespacio, en colaboración con otras autoridades, desde el ámbito de su competencia, así como el empleo de la inteligencia artificial como herramienta tecnológica" (énfasis propio). Para tal efecto, el artículo 5 establece en sus niveles de mando las Unidades de Soporte Estratégico para la Ciberdefensa e Inteligencia Artificial, definidas en el artículo 37 de la ley como: "unidades operativas especializadas en la aplicación de capacidades tecnológicas institucionales y soluciones basadas en inteligencia artificial como herramienta analítica para optimizar los procesos y efectividad en el desarrollo de las operaciones en el ámbito del ciberespacio, uso del espectro electromagnético, operaciones de información y dominio espacial, en apoyo a las operaciones navales" (énfasis propio).

- En noviembre de 2025, la Secretaría de Seguridad Ciudadana anunció que la Unidad de Inteligencia Financiera (UIF) de la Secretaría de Hacienda y Crédito Público entró en una "nueva etapa de prevención" al informar que la UIF y las dependencias de seguridad "desarrollarán mecanismos de detección temprana, modelos predictivos basados en inteligencia artificial y nuevas tipologías sectoriales para identificar conductas inusuales antes de que generen un daño al sistema financiero".[4]

Las facultades y medidas orientadas a emplear herramientas de inteligencia artificial por parte de autoridades de seguridad y defensa son particularmente preocupantes en el contexto de las reformas aprobadas en julio de 2025, las cuales crean una infraestructura de vigilancia masiva en México como resultado de la creación de una identificación biométrica obligatoria para toda la población. La identidad biométrica será exigida para el acceso a servicios públicos y privados. De este modo, la actividad derivada del uso de esta identificación se registrará en tiempo real en la nueva "Plataforma Central de Identidad" regulada por la Ley General de Población y la Ley General en materia de desaparición forzada.

Todo lo anterior se traduce en una concentración de bases de datos públicas con información biométrica y personal la cual será objeto de tratamiento y procesamiento predictivo por parte de autoridades a través de herramientas de inteligencia artificial.

4. Quadratín México, "Así operaban los 13 casinos investigados por lavado de dinero". 12 de noviembre de 2025. Disponible en: https://mexico.quadratin.com.mx/asi-operaban-los-13-casinos-investigados-por-lavado-de-dinero/

b. Sistemas predictivos en el acceso a programas sociales

El repositorio de Algoritmos Públicos del Centro de Investigación y Docencia Económicas (CIDE) contiene resultados de múltiples solicitudes de información realizadas a distintas entidades gubernamentales sobre el uso de sistemas automatizados en el ejercicio de sus funciones.[5] La Secretaría Ejecutiva del Sistema Nacional Anticorrupción (SESNA) informó que se encuentra en desarrollo un proyecto relacionado con el uso de algoritmos para programas sociales. SESNA establece diversas prioridades dentro de la Política Nacional Anticorrupción (PNA), en las que se incluyen el impulso al uso de tecnologías avanzadas, el análisis de datos masivos y la inteligencia artificial, con el fin de identificar riesgos y mejorar la gestión, la auditoría y la fiscalización estratégica en el sector público. Hasta el momento, no hay más información técnica disponible públicamente sobre el proyecto.

c. Sistemas predictivos en materia fiscal

En 2024, el Servicio de Administración Tributaria anunció en el Plan Maestro 2024[6] el uso de inteligencia artificial para una mejor planeación en los procesos de recaudación. El objetivo es implementar modelos de analítica de grafos y *machine learning* para la clasificación de los contribuyentes de riesgo, la identificación de redes complejas de elusión y evasión, así como la detección de inconsistencias en comprobantes fiscales digitales (CFDI) asociadas con el contrabando y empresas fachada.

Para 2026, el SAT anunció que su plataforma de inteligencia artificial "ampliará su alcance más allá de la banca tradicional, incorporando al análisis a las fintech y plataformas digitales que operan en el país", con lo que movimientos en servicios como *PayPal* o *Mercado Pago* también serán auditables. El sistema de IA del SAT detectará como acciones irregulares: a) depósitos en efectivo recurrentes; b) ingresos derivados por ventas por catálogo; c) préstamos entre familiares sin respaldo legal; y d) diferencias entre gastos con tarjetas de crédito e ingresos declarados.[7] Si bien algunos analistas consideran que la implementación de estos sistemas automatizados ha ayudado a incrementar la recaudación fiscal,[8] otros advierten que la falta

---

5. Algoritmos CIDE, Repositorio - Algoritmos CIDE, Centro de Investigación y Docencia Económicas (CIDE), última consulta 2 de julio de 2025, https://algoritmoscide.org/repositorio/

6. Servicio de Administración Tributario. *Plan Maestro 2024*. Gobierno de México. 22 de enero de 2024. Disponible en: https://www.gob.mx/sat/documentos/plan-maestro-2024-sat

7. El Cronista México. "El SAT usará inteligencia artificial para fiscalizar todas las cuentas bancarias del país: habrá control estricto para estos mexicanos". Comunicado oficial. 2 de febrero de 2026. Disponible en: https://www.cronista.com/mexico/finanzas-economia/el-sat-usara-inteligencia-artificial-para-fiscalizar-todas-las-cuentas-bancarias-del-pais-habra-control-estricto-para-estos-mexicanos/

8. Canabal, P. "La Silla Rota: SAT utiliza IA para fiscalizar contribuyentes y aumenta recaudación". *Bakertilly*. 25 de mayo de 2025. Disponible en: https://www.bakertilly.mx/noticias/sat-utiliza-ia-para-fiscalizar-contribuyentes-y-aumenta-recaudaci%C3%B3n

de una regulación específica sobre el uso de sistemas de inteligencia artificial en México, así como la opacidad algorítmica sobre la clasificación de "contribuyentes de riesgo", pueden generar exclusión y discriminación, entre otras violaciones a derechos humanos.[9] R3D coincide y considera que la falta de transparencia en la adopción y evaluación de los algoritmos utilizados para fines de recaudación fiscal genera riesgos e incertidumbre legal para usuarios del sistema financiero que pudieran ser víctimas de una clasificación equivocada. Hasta ahora se desconoce si existen mecanismos de queja y remediación por errores en la clasificación arrojada por los algoritmos.

d. Sistemas automatizados y derechos laborales en México

La Ley Federal del Trabajo, reformada en diciembre de 2024 establece que las reglas para la asignación de servicios o tareas a través de algoritmos deben ser transparentes, claras y conocidas por los trabajadores de plataformas digitales. También se establece un mecanismo de transparencia social por el cual requiere que las plataformas elaboren un documento de política de gestión algorítmica del trabajo. El aspecto relevante de estas normas es que las reglas de asignación deben explicar de manera clara y accesible las consecuencias del cumplimiento de instrucciones, el impacto de calificaciones de terceros, los incentivos y las penalizaciones, la existencia de categorías que influyan en la asignación de tareas y otros criterios relevantes. Este reporte deberá integrarse al contrato laboral y ser conocido desde el inicio de la relación o ante cualquier modificación. Además, se establece que los algoritmos deben ser razonables, no poner en riesgo la salud ni la integridad del trabajador ni generar discriminación. De igual forma, en el artículo 291-P, establecen un mecanismo de apelación y rendición de cuentas para los trabajadores que debe ser gestionado por una persona y no por un sistema automatizado.

Desde R3D enfatizamos que la regulación tiene limitaciones importantes que requieren cambios para garantizar los derechos involucrados en el uso de estos sistemas automatizados. Al menos, contar con: **evaluaciones de impacto** de los sistemas automatizados usados por las plataformas digitales; procedimientos de **revisión y supervisión humana calificada** en las decisiones automatizadas; **evaluaciones periódicas a los sistemas automatizados** por personas capacitadas e imparciales; la obligación de **informar a las personas trabajadoras** sobre los resultados de estas evaluaciones; acciones cuando los sistemas evaluados representan un riesgo o lesionen los derechos de las personas trabajadoras (incluyendo **modificar el sistema o suspender su uso**); la obligación de que toda decisión relacionada con suspensiones o la finalización de la relación laboral **sea tomada por una persona** y no por un sistema; evaluaciones

9. Palacios, S. "¿Es legal que el SAT aplique IA para clasificar contribuyentes?". *Alto Nivel*. 26 de febrero de 2025. Disponible en: https://www.altonivel.com.mx/es-legal-que-el-sat-aplique-ia-para-clasificar-contribuyentes/

de riesgos de los sistemas puedan en la seguridad y salud de los trabajadores; garantizar acceso a **mecanismos de resolución de disputas, reparación y compensación**.[10]

II. Participación social, transparencia y acceso a la información

En México, el diseño, el desarrollo, la evaluación y la implementación de sistemas de inteligencia artificial, así como su adquisición, se mantienen sin una participación sustantiva de sociedad civil especializada en las implicaciones de la inteligencia artificial en los derechos humanos. De manera generalizada, las decisiones estatales sobre el uso y despliegue de estas herramientas suelen centralizarse en un par de instituciones (tal como sucede con la Agencia de Transformación Digital o las autoridades de seguridad pública), que hasta ahora no han cumplido con sus obligaciones en materia de de transparencia proactiva ni han desarrollado mecanismos de participación pública conforme a los cuales tomen decisiones informadas que reconozcan y consideren los derechos humanos involucrados en el diseño, desarrollo, uso, evaluación e implementación de sistemas de inteligencia artificial por parte de autoridades y actores privados.

Los mecanismos para compartir información se han centrado de manera sistemática en las conferencias matutinas del gobierno federal en donde dan noticias relacionadas con el desarrollo de políticas y regulaciones relacionadas con la IA, pero la gran mayoría de estas decisiones se toman sin contar con insumos y participación de la sociedad civil organizada. Lo que suele presentarse es el estado final del proyecto para ser aprobado legislativamente o para ser implementado por la autoridad correspondiente.

Un gran ejemplo sobre estas prácticas relativas al uso de sistemas de inteligencia artificial que excluyen las perspectivas de sociedad civil es el paquete legislativo en materia de seguridad aprobado a mediados de 2025 y mencionado en la sección anterior. Dentro de este paquete, la Ley de Inteligencia establece que toda la información relacionada con las acciones de investigación y vigilancia deberá quedar reservada y que sobre esta información podrán utilizarse sistemas de IA para su procesamiento, análisis e incluso la persecución de delitos. Pese a los intentos por contribuir en el diseño y elaboración de las reformas como organización de la sociedad civil experta, el paquete legislativo se aprobó en los términos que fue presentado por las autoridades. Lo cual, de la mano de la clasificación de las acciones del Estado como reservadas, hará muy difícil conocer cómo, de qué manera, bajo qué parámetros y con qué salvaguardas para los derechos estos sistemas serán utilizados. R3D y otras organizaciones de sociedad civil planteamos de manera contundente la necesidad de tomar en consideración la evidencia de

10. R3D. "Reforma laboral para trabajadores de plataformas es un avane". Disponible en: https://r3d.mx/2024/12/10/la-reforma-laboral-para-trabajadores-de-plataformas-digitales-es-gran-avance-pero-debe-garantizar-la-proteccion-de-datos-y-la-transparencia-algoritmica/

que en México existen prácticas generalizadas de vigilancia ilegal por parte distintas instituciones estatales y militares, en violación de los derechos de personas defensoras de derechos humanos, periodistas, activistas y opositores políticos. Nuestra contribución buscaba generar una legislación que impusiera controles democráticos en el uso de sistemas de Inteligencia Artificial. Sin embargo, la propuesta original fue aprobada sin contrapesos políticos o de instituciones públicas de derechos humanos.[11]

La situación de opacidad debe leerse a la luz del proceso de desmantelamiento institucional que se vivió en el último año en el país. El gobierno desapareció y desarticuló el sistema autónomo y constitucional de transparencia que existía, creando un sistema de acceso a la información pública centralizado en el poder Ejecutivo, lo cual limita la posibilidad de acceso a la información de interés público relacionada con la gestión del gobierno, particularmente la relativa al uso y desarrollo de sistemas de inteligencia artificial para el quehacer público.

Sucede lo mismo con la desaparición de la independencia del Poder Judicial y con el proceso de reestructuración y de elección popular de juezas y jueces. Con él, la integración completa de la Suprema Corte de Justicia de la Nación, el tribunal constitucional mexicano, fue modificada para que perfiles impulsados por el gobierno ocuparan los cargos, afectando fuertemente la independencia del Poder Judicial. Con ambos embates, es posible anticipar que las violaciones a derechos humanos derivadas del uso de sistemas de IA no serán revisadas por organismos independientes ni autónomos, ni contarán con las garantías judiciales necesarias para proteger los derechos individuales y colectivos de las personas.

III. Desarrollo, provisión y uso de inteligencia artificial por parte de actores privados

Las principales empresas que operan en México utilizando distintos tipos de sistemas de IA e IA generativa son: Meta, Google, OpenAI, X, Amazon, IBM, Microsoft, Xira, Rappi, Uber, entre otras. Los sistemas que utilizan se dirigen principalmente al desarrollo de productos específicos de generación de texto, imagen o video, pero también al desarrollo de sistemas de recomendación en sus sistemas de moderación de contenidos y al despliegue de publicidad y de monetización.

En la actualidad, no existe un marco regulatorio que dé mecanismos de transparencia, debida diligencia y rendición de cuentas dirigidos a la contratación pública de servicios y/o productos de estas empresas, pero tampoco un sentido más amplio sobre el cumplimiento de obligacio-

11. R3D. "El Estado de la Vigilancia", Ciudad de México, 2025. Disponible en: https://r3d.mx/wp-content/uploads/EDLV_2025.pdf

nes en materia de respeto a los derechos humanos en el desarrollo de dichos sistemas (mucho menos en cuanto a la supervisión para el cumplimiento de dichas obligaciones).

No hay evaluaciones públicas ni privadas que sean públicas, abiertas ni participativas para garantizar que las decisiones relativas al diseño, desarrollo, entrenamiento, implementación y uso de inteligencia artificial estén alineadas con el deber de respeto de los derechos humanos ni con los estándares interamericanos sobre empresas y derechos humanos. Por el contrario, ante peticiones y llamados a transparentar sus prácticas y observar sus responsabilidades en materia de derechos humanos, las empresas normalmente argumentan limitaciones en materia de propiedad intelectual y competencia para no cumplir con obligaciones de transparencia significativa (incluida la transparencia algorítmica). En los pocos casos en los que se incluye a la sociedad civil y a organizaciones especializadas de derechos humanos para evaluar algún producto o servicio, la proporción de información que se revela y las exigencias que se piden para participar en dichos encuentros (sobre acuerdos de confidencialidad, por ejemplo), hacen que la posibilidad de impacto y supervisión sea meramente testimonial y no sustantiva.

Cabe recalcar que desde la alineación política de estas empresas de origen estadounidense con la nueva administración de los Estados Unidos de América en 2025, sus prácticas en materia de involucramiento y consulta con la sociedad civil se han visto significativamente reducidas. Las empresas han reconocido expresamente que la sociedad civil dejó de ser una parte interesada prioritaria para el desarrollo y evaluación de sus políticas y servicios.

Respecto a las evaluaciones de riesgo sobre los derechos humanos, no existe evidencia alguna de que las empresas que desarrollan productos y brindan servicios digitales basados en sistemas de inteligencia artificial en México realicen evaluaciones que estén abiertas a la auditoría y seguimiento independientes por parte de la sociedad civil o de otro sujeto independiente (mucho menos de instituciones públicas). En las ocasiones en que R3D solicitó involucrarse de manera sustantiva para participar en la evaluación de riesgos de productos y servicios de las empresas relacionados con investigaciones sobre censura electoral en plataformas digitales, evaluación y seguimiento de sistemas de moderación de contenidos, transparencia algorítmica y garantías de protección de datos y privacidad, por mencionar algunas, las empresas se rehusaron a proporcionar la información relevante. Comúnmente, sus respuestas se refirieron a la información general agregada de sus reportes generales o anuales de derechos humanos. Respecto de las auditorías, utilizan los mismos argumentos relativos a información protegida por el derecho de propiedad intelectual o a la afectación de sus intereses económicos al implicar una desventaja en términos de competencia frente a otros sujetos.

IV. Impacto del uso de inteligencia artificial, incluyendo la inteligencia artificial generativa, en derechos específicos

Las tecnologías de inteligencia artificial, en especial de carácter generativo (texto, imagen, video), afectan de distintas maneras el derecho a la libertad de expresión en alguna de sus dimensiones (individual, social y democrática). El despliegue de la IA generativa se ha hecho predominantemente siguiendo un modelo de negocios que es incompatible con el deber de respeto a los derechos humanos que las empresas tienen, pero que también afecta a otros deberes como el de transparencia (en especial por lo que corresponde a la transparencia algorítmica) y a la diligencia debida. Hay al menos **tres** formas claras en que estos sistemas afectan este derecho y otros derechos relacionados:

1. El uso de la IA para generar desinformación y dañar el ecosistema de información en general y el sistema de conocimiento en particular

El uso de la IA para generar desinformación en México no es nuevo y existen numerosos casos en los que ha tenido lugar, principalmente en contextos electorales o relacionados con el ejercicio del poder político y gubernamental. En las últimas elecciones presidenciales de 2024, se documentó el uso de sistemas de IA por partidos políticos para llevar a cabo publicidad dirigida, analizar el comportamiento de votantes, crear contenidos específicos para dirigirlos a grupos demográficos específicos, y generar contenidos para desinformar abiertamente.

Uno de los ejemplos más claros fue el del video *deepfake* creado con IA sobre la entonces candidata Claudia Sheinbaum, en el que aparecía supuestamente convocando a la población a invertir en una plataforma financiera fraudulenta. Pero ese no fue el único video de desinformación generado por IA contra la entonces candidata, sino que también circuló otro sobre una supuesta propuesta de que cerraría las iglesias al asumir la presidencia del país, con la intención de afectar al electorado católico en México.[12] En la desinformación sobre laahora presidenta de México se repetían estereotipos de carácter sexista que la relacionaban con no ser apta para este cargo por ser mujer, la menospreciaban por motivos

12. Barenque, Andrea. "Addressing Disinformation in a Politically Polarized Landscape", The Blue Owl Group, agosto de 2024. https://www.blueowlgrp.com/er24-mexico; Lagos, Anna. "Elecciones México 2024: Desinformación, IA y el reto de la democracia en un país polarizado", WIRED, 20 de agosto de 2024. Disponible en: https://es.wired.com/articulos/elecciones-mexico-2024-desinformacion-ia-y-el-reto-de-la-democracia-en-un-pais-polarizado; Lagos, Ana. "Claudia Sheinbaum, víctima de *deepfake*: su imagen fue utilizada en un video fraudulento generado por IA", WIRED, 25 de enero de 2024. Disponible en: https://es.wired.com/articulos/claudia-sheinbaum-victima-de-deepfake-su-imagen-fue-utilizada-en-un-video-fraudulento-generado-por-ia; Melina Barbosa. "Circulan video montajes de Claudia Sheinbaum diciendo que cerrará iglesias cuando gane la presidencia", verificado, 29 de mayo de 2024. Disponible en: https://verificado.com.mx/claudia-sheinbaum-cierre-de-iglesias-presidencia/#google_vignette

religiosos e incluso la presentaban como una supuesta comunista que estaba en contra de la propiedad privada.[13]

La candidata Xóchitl Gálvez también resultó afectada de manera similar, a partir de tecnologías de IA y basándose en estereotipos de género para aludir a una supuesta menor capacidad de las mujeres para participar en política. En su caso, se difundieron al menos dos videos que fueron generados o alterados con tecnologías de inteligencia artificial. El primero, donde supuestamente aparecía ondeando una bandera al revés, en el que se le criticaba por su ignorancia y sus capacidades. En otro video aparecía supuestamente diciendo que eliminaría los programas sociales existentes del gobierno, en una supuesta postura antipopulista, pero en ese caso se trataba de videos editados y complementados de manera que parecía decir cosas que nunca afirmó.[14] En este contexto, distintas y distintos especialistas coincidieron en que los ataques de desinformación dirigidos a las candidatas mujeres destacaban por sus contenidos sexistas, a diferencia de aquella desinformación que circuló contra sus pares hombres.[15]

La IA generativa se utiliza cada vez con mayor frecuencia para desinformar y manipular a las personas. **Dichas tecnologías están diseñadas para privilegiar la generación de contenidos (imágenes, texto, video) que respondan a las peticiones de sus usuarios y que les complazcan, incluso si esto afecta la precisión y veracidad de la información que se produce**. Este problema no es necesariamente intrínseco, sino que el desarrollo de la IA se ha hecho de un modo en el que la preocupación por la información y su conexión con la verdad o con los hechos empíricos no es relevante. Lo anterior facilita el desprendimiento o ruptura con la realidad y la posibilidad de generar "realidades" compartidas por grupos de

---

13. Klepper, David.. "Los tópicos sexistas y la desinformación campan por Internet mientras México se alista para votar", Los Angeles Times, 1 de junio de 2024. Disponible en: https://www.latimes.com/espanol/mexico/articulo/2024-06-01/los-topicos-sexistas-y-la-desinformacion-campan-por-internet-mientras-mexico-se-alista-para-votar; García, Syndy. "Elecciones presidenciales en México: Las desinformaciones y los ataques en redes del proceso 2024", Voz de América, 30 de mayo de 2024. https://www.vozdeamerica.com/a/elecciones-presidenciales-en-mexico-las-desinformaciones-y-los-ataques-en-redes-del-proceso-2024/7636345.html

14. Soto, Diana. 2024. "¡Falso! Xóchitl no ondeó la bandera de México con el escudo al revés", El Sabueso (Animal Político), 15 de mayo de 2024. Disponible en: https://animalpolitico.com/verificacion-de-hechos/desinformacion/xochitl-no-ondeo-bandera-invertida; AFP. "Fake news que circularon sobre las elecciones 2024 en México", El Economista, 01 de junio de 2024. Disponible en: https://www.eleconomista.com.mx/politica/Fake-news-que-circulan-sobre-las-elecciones-2024-en-Mexico-20240531-0086.html

15. Klepper, David.. "Los tópicos sexistas y la desinformación campan por Internet mientras México se alista para votar", Los Angeles Times, 1 de junio de 2024. Disponible en: https://www.latimes.com/espanol/mexico/articulo/2024-06-01/los-topicos-sexistas-y-la-desinformacion-campan-por-internet-mientras-mexico-se-alista-para-votar; VerificaRTVE. 2024. "La desinformación en México durante la campaña de las elecciones presidenciales", Corporación de Radio y Televisión Española, 29 de mayo de 2024. Disponible en: https://www.rtve.es/noticias/20240529/desinformacion-mexico-campana-elecciones-presidenciales/16123707.shtml

usuarios sin que exista una en común (no sólo de forma malintencionada sino como resultado del diseño propio).[16]

Si bien la tecnología como tal ha resultado útil en otros contextos, como en algunos casos relacionados con entornos laborales y académicos, existe un serio riesgo en su uso cuando los modelos de lenguaje de gran tamaño facilitan la creación de textos inverosímiles solicitados con la intención de desinformar o cuando, por su propia mecánica y diseño, generan información fabricada que no se basa en la realidad pero responde a la petición de la persona usuaria. Estos grandes modelos de lenguaje no han sido entrenados para proveer información veraz ni con la cual contrastar sus resultados. La tendencia a no revisar, contrastar o cuestionar la información producida por IA pone en riesgo el sistema de conocimiento y daña las opciones de discernimiento para todas las personas.[17]

En este contexto, también **encontramos preocupante que**, incrementalmente, **la primera línea de entrada para acceder y consultar información en la red es el servicio de estos sistemas de inteligencia artificial generativa**. Es decir, las personas recurren de manera sistemática a estos sistemas para informarse, ya que parecen arrojar resultados de manera confiable, similar a la práctica de buscar información en motores de búsqueda. Todo esto sin advertir las limitaciones de diseño de esas tecnologías ni la forma en que generan sus respuestas, y sin dimensionar también la forma en que estos sistemas fallan e inventan contenidos de forma cotidiana (sin mencionar los numerosos casos en los que los sistemas de IA simplemente mienten a las personas usuarias o utilizan la información proporcionada por ellas para producir información que puede dañarles o incluso causarles la muerte).[18]

Pero el impacto en el ecosistema de información en general no es el único nivel en el que estas tecnologías afectan las bases fácticas de la información que son necesarias para que las personas puedan formar sus opiniones, desarrollar sus creencias y tomar decisiones en la vida democrática. De manera creciente, la IA se está utilizando para crear lo que ha sido denominado "contenido bazofia generado por IA" o "*AI Slop*", el cual se refiere a contenido sintético generado para ser llamativo y repetitivo sin mucho valor para la construcción crítica de la información. Sectores académicos alertan que estos contenidos han dado pie a investigaciones sin metodologías rigurosas,

---

16. Mayer, Macey. *AI wants to make you happy. Even if it has to bend the truth*, CNET, Nov. 16, 2025. Disponible en: https://www.cnet.com/tech/services-and-software/ai-wants-to-make-you-happy-even-if-it-has-to-bend-the-truth/

17. van Rooij, Iris. *AI Slop and the destruction of knowledge*, 2025. Disponible en: https://irisvanrooijcogsci.com/2025/08/12/ai-slop-and-the-destruction-of-knowledge/

18. Mayer, Macey. *AI wants to make you happy. Even if it has to bend the truth*, CNET, Nov. 16, 2025. Disponible en: https://www.cnet.com/tech/services-and-software/ai-wants-to-make-you-happy-even-if-it-has-to-bend-the-truth/

impactando el sistema científico y la academia, así como la infraestructura del conocimiento, a estudiantes, profesores, y al público que puede llegar directa o indirectamente a esa información.[19]

En términos de los impactos en la dimensión colectiva o social del derecho a la libertad de expresión, la desinformación habilitada por sistemas de inteligencia artificial generativa interfiere y genera obstáculos para recibir y conocer libremente informaciones, opiniones, relatos y noticias sobre hechos y la vida pública; tienen la capacidad de distorsionar la recepción de dicha información en tanto es cada vez más difícil distinguir los contenidos reales de aquellos que son producidos por la IA generativa que se usan de manera intencionada para manipular la opinión de las personas.

La dimensión individual de la libertad de expresión se ve afectada por una tendencia conforme a la cual las tecnologías de IA permiten o incluso promueven la producción de lo que se ha denominado "contenido sintético perjudicial a escala", que afecta directamente los derechos de las personas y se expande a daños en su intimidad o su autonomía, sirviendo como un obstáculo para ejercer su derecho a la libertad de expresión y en muchos casos convirtiéndolas en víctimas de violencia en línea. El ejemplo más claro se da con la creación, almacenamiento y distribución de contenidos de abuso sexual infantil (CASI) y con la creación y difusión sin consentimiento de imágenes pornográficas (fenómenos que afectan de manera diferenciada a mujeres y niñas).[20]

Así, uno de los usos más comunes de la IA generativa es el de ser producida para ridiculizar, humillar, acosar y violentar a mujeres y niñas a partir de medidas de exposición, extorsión y chantaje, pero también el de generar audio y video con el ánimo de engañar a la gente o dañar de manera específica a grupos en posición de vulnerabilidad.

El aumento de incidentes y riesgos registrados por el "AI Incidents and Hazards Monitor" de la OCDE, el cual monitorea incidentes de "riesgo" generados por IA en los países de la OCDE, registró hasta 250 por mes en octubre de 2025, cuando registraba menos de 25 casos mensuales en el 2022.[21] Sobre México, se han registrado 64 incidentes en los últimos seis años. El más reciente pudo observarse tras el abatimiento de uno de los líderes de un grupo del narcotráfico. En este caso, las redes se inundaron de noticias falsas producidas por tecnologías de inteligencia artificial que buscaban generar terror en la población y mostraban hechos de violencia que no sucedieron en la realidad. De acuerdo a varias personas expertas, muchas de estas noticias fue-

---

19. van Rooij, Iris. *AI Slop and the destruction of knowledge*, 2025. Disponible en: https://irisvanrooijcogsci.com/2025/08/12/ai-slop-and-the-destruction-of-knowledge/

20. Bengio, Yoshua *et al*. "Informe Internacional sobre la Seguridad de la IA", *Mila - Quebec AI Institute*, febrero de 2026 (DSIT 2026/001, 2026), pp. 13, 50, 53. Disponible en: https://internationalaisafetyreport.org

ron generadas por miembros de los cárteles y por nuevas personalidades denominadas "narcoinfluencers" (personas mediáticas en redes sociales que glorifican o apoyan a estos grupos).[22]

En este contexto, R3D coincide con el *Informe Internacional sobre la Seguridad de la IA*, en donde señala de manera contundente que uno de los riesgos sistémicos de este tipo de tecnologías es el que se genera a los entornos de información por la producción de desinformación y la forma en que ésta degrada la información disponible de las personas (creando *entornos informativos de baja calidad o entornos sesgados*), e influye en la formación de la opinión y las creencias de las mismas, afectando su autonomía para razonar de manera informada y tomar decisiones autónomas.[24] Así:

> "En muchos casos, la IA ya está transformando la forma en que las personas acceden a la información, toman decisiones y resuelven problemas, con aplicaciones en sectores que van desde el desarrollo de software a los servicios jurídicos o la investigación científica".[25]

La capacidad de manipular y desinformar con información generada por estas tecnologías aumenta con la escala de los propios modelos, por lo que a mayor poder de cómputo, mayor es la probabilidad de que logre modificar las creencias de las personas o de generar desinformación.[26] Esto introduce una variable fundamental para el respeto a los derechos humanos y la

---

21. H. Ajder, G. Patrini, F. Cavalli, L. Cullen, "The State of Deepfakes: Landscape, Threats, and Impact" (Deeptrace, 2019); https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

22. Reuters. *Noticias falsas, narcoinfluencers e IA amplificaron la violencia y el miedo tras abatimiento de "El Mencho"*, La Jornada, 25 de febrero de 2026. Disponible en: https://www.jornada.com.mx/noticia/2026/02/25/politica/noticias-falsas-narcoinfluencers-e-ia-amplificaron-la-violencia-y-el-miedo-tras-abatimiento-de-el-mencho

23. Bengio, Yoshua *et al*. "Informe Internacional sobre la Seguridad de la IA", *Mila - Quebec AI Institute*, febrero de 2026 (DSIT 2026/001, 2026), p. 56. En muchos casos, es posible advertir la existencia de una "deferencia de credibilidad" por la que las personas usuarias de estas tecnologías confían en los contenidos generados por estos sistemas sin revisarlos, verificarlos o cuestionarlos.

24. L. Malmqvist, "Sycophancy in Large Language Models: Causes and Mitigations" in Lecture Notes in Networks and Systems (Springer Nature Switzerland, Cham, 2025), pp. 61–74: https://doi.org/10.1007/978-3-031-92611-2_5; E. Perez, S. Ringer, K. Lukosiute, K. Nguyen, E. Chen, S. Heiner, C. Pettit, C. Olsson, S. Kundu, S. Kadavath, A. Jones, A. Chen, B. Mann, B. Israel, B. Seethor, C. McKinnon, C. Olah, … J. Kaplan, "Discovering Language Model Behaviors with ModelWritten Evaluations" in Findings of the Association for Computational Linguistics: ACL 2023, A. Rogers, J. BoydGraber, N. Okazaki, Eds. (Association for Computational Linguistics, Toronto, Canada, 2023), pp. 13387–13434; https://doi.org/10.18653/v1/2023.findings-acl.847; L. Ranaldi, G. Pucci, When Large Language Models Contradict Humans? Large Language Models' Sycophantic Behaviour, arXiv [cs.CL] (2025); http://arxiv.org/abs/2311.09410

25. Bengio, Yoshua *et al*. "Informe Internacional sobre la Seguridad de la IA", Mila - Quebec AI Institute, febrero de 2026 (DSIT 2026/001, 2026), p. 15.

26. K. Hackenburg, B. M. Tappin, L. Hewitt, E. Saunders, S. Black, H. Lin, C. Fist, H. Margetts, D. G. Rand, C. Summerfield, The Levers of Political Persuasion with Conversational AI, arXiv [cs.CL] (2025): http://arxiv.org/abs/2507.13919

estabilidad de las democracias, dado que da un mayor poder de control a las grandes empresas tecnológicas que tengan intereses económicos que se beneficien de conflictos políticos o institucionales en un país (tal como se ha mostrado en distintas ocasiones con el uso político de la red "X", de Elon Musk, en casos como el de EUA, Brasil o Alemania).[27]

2. El uso de la IA y su relación con la libertad de opinión, pensamiento y autonomía para la deliberación democrática

Otro de los impactos documentados en los que las tecnologías de IA afectan el derecho a la libertad de expresión, el derecho a la autonomía y su relación con la libertad de pensamiento se observan en dos fenómenos concretos: el del **daño cognitivo** generado por el uso de la IA y el del "**sesgo de automatización**" que existe en sus personas usuarias (no sólo las personas en general sino en ámbitos especializados como el de la medicina, la programación o la física).[28]

Cada vez es más común el uso cotidiano de la IA para la búsqueda de información, la recomendación de contenidos, el acceso a servicios, el uso de redes sociales o el uso de asistentes de chat que proporcionan información y, en muchos casos, influencian decisiones de las personas. La IA generativa produce "contenido persuasivo" que tiene la capacidad de influenciar en la construcción individual y colectiva de sistemas de creencias, razonamientos y opiniones propias de cada persona, lo cual puede generar una *deferencia* por esa información dada por el sistema de IA. En muchos casos, el impacto de estas tecnologías no sólo *erosiona la confianza* de las personas sobre el propio sistema de información y su capacidad de interactuar individualmente con él, sino que puede llevar a situaciones de disminución de la autonomía al interferir con la libertad de pensamiento y el proceso de toma de decisiones. Entre los casos extremos documentados a nivel global se encuentran aquellos relacionados con la creación de dependencia emocional y psicológica de las tecnologías, de realización de acciones peligrosas o del reforzamiento de creencias dañinas o violentas[29] (como

---

27. AP. "¿Cómo Elon Musk utiliza la "libre expresión" de X para difundir cuestionadas posiciones al mundo?", *France 24*, 13 de agosto de 2024. Disponible en: https://www.france24.com/es/ee-uu-y-canad%C3%A1/20240813-c%C3%B3mo-elon-musk-utiliza-la-libre-expresi%C3%B3n-de-x-para-difundir-cuestionadas-posiciones-al-mundo; Limón, Raúl. "La red X, de Elon Musk, es una plataforma para el "abuso político" al relegar a los moderados y tratarlos como enemigos", *EL PAÍS*, 14 de noviembre de 2024. Disponible en: https://elpais.com/tecnologia/2024-11-14/la-red-x-de-elon-musk-es-una-plataforma-para-el-abuso-politico-al-relegar-a-los-moderados-y-tratarlos-como-enemigos.html

28. Bengio, Yoshua *et al*. "Informe Internacional sobre la Seguridad de la IA", *Mila - Quebec AI Institute*, febrero de 2026 (DSIT 2026/001, 2026), pp. 13-14, 56-57.

29. Existe evidencia del aumento del consumo de los llamados "acompañantes de IA", con un consumo de decenas de millones de personas y con resultados en distintos casos de dependencia emocional, delirios y incluso llegar a quitarse la vida tras interacciones con chatbots. Ver: A. R. Liu, P. Pataranutaporn, P. Maes, Chatbot Companionship: A Mixed-Methods Study of Companion Chatbot Usage Patterns and Their Relationship to Loneliness in Active Users, arXiv [cs.HC] (2025); http://arxiv.org/abs/2410.21596; Z. Qian, M. Izumikawa, F. Lodge, A. Leone, Mapping the Parasocial AI Market: User Trends, Engagement and Risks, arXiv [cs.CY] (2025); http://arxiv.org/abs/2507.14226; J.

sucede con los sistemas que privilegian discursos radicalizados que generan odio interpersonal y que pueden incitar a la violencia o generar adicción de uso sobre las personas usuarias).[30]

Si bien no existe una definición jurídica sobre lo que constituye el "pensamiento" en el contexto de la libertad de pensamiento y consciencia en el derecho internacional de los derechos humanos, sí existen atributos claros de este derecho protegido por el artículo 18 del Pacto Internacional de Derechos Civiles y Políticos, que a diferencia de la CADH, incluye la libertad de pensamiento como elemento autónomo de la libertad de conciencia y religión de las personas. La libertad de pensamiento no aplica únicamente en relación con la libertad de pensamiento religioso, el Comité de Derechos Humanos de Naciones Unidas determinó que abarca **libertad de pensamiento en todas las cuestiones y convicciones personales**.[31]

El Relator Especial sobre libertad de religión o de creencias de Naciones Unidas ha establecido que "pese a que el pensamiento y la expresión son distintos desde el punto de vista conceptual y práctico, se encuentran inmersos en un bucle perpetuo de influencia mutua en que la expresión sirve de vehículo para intercambiar y desarrollar pensamientos, y los pensamientos sirven de alimento a la expresión."[32] Respecto a la relación del pensamiento con la opinión, el Relator respalda el planteamiento conforme al cual la libertad de opinión está estrechamente relacionada con la libertad de pensamiento dentro del fuero interno, y este proceso interno entre pensamiento y opinión interactúa con el externo que se traduce en la expresión. Entonces, se podría

---

De Freitas, N. Castelo, A. K. Uğuralp, Z. OğuzUğuralp, Lessons from an App Update at Replika AI: Identity Discontinuity in Human-AI Relationships, arXiv [cs.HC] (2024); http://arxiv.org/abs/2412.14190; H. Bai, J. G. Voelkel, S. Muldowney, J. C. Eichstaedt, R. Willer, LLM-Generated Messages Can Persuade Humans on Policy Issues. Nature Communications 16, 6037 (2025); https://doi.org/10.1038/s41467-025-61345-5; V. Bakir, A. McStay, Move Fast and Break People? Ethics, Companion Apps, and the Case of Character. ai. AI & Society (2025); https://doi.org/10.1007/s00146-025-02408-5; B. P. Billauer, Murder without Redress - the Need for New Legal Solutions in the Age of Character -AI (C.a.i.) (2025); https://doi.org/10.2139/ssrn.5107942

30. A inicios de este año, la Comisión Europea de Derechos Humanos determinó, en una investigación preliminar, la responsabilidad de la plataforma digital *TikTok* por el "diseño adictivo" de su aplicación para generar daños físicos y mentales en sus usuarios, incluidos menores y personas adultas vulnerables, y en violación de la Ley de Servicios Digitales. En: European Commission. "Commission preliminarily finds TikTok's addictive design in breach of the Digital Services Act", Feb 5, 2026. Disponible en: https://ec.europa.eu/commission/presscorner/detail/en/ip_26_312

31. Observación General No. 22, Comentarios generales adoptados por el Comité de los Derechos Humanos, Artículo 18 - Libertad de pensamiento, de conciencia y de religión, párr. 1, 48º período de sesiones, U.N. Doc. HRI/GEN/1/Rev.7 at 179 (1993); Office for Democratic Institutions and Human Rights (ODIHR). *Think Again: Freedom of Thought in the Age of Artificial Intelligence*, OSCE, 2025. Disponible en: https://odihr.osce.org/sites/default/files/f/documents/0/e/597450.pdf

32. Asamblea General de la Organización de las Naciones Unidas. "Informe provisional del Relator Especial sobre la libertad de religión o de creencias, Ahmed Shaheed", párr. 18, 5 de octubre de 2021, A/76/380. Disponible en: https://docs.un.org/es/A/76/380

decir que el pensamiento es un proceso y la opinión el resultado de dicho proceso.[33] Bajo este marco analítico, el Relator propuso un esquema de atributos de la libertad de pensamiento, los cuales son relevantes en el contexto de impactos cognitivos e influencia de la IA en el proceso de creación de pensamientos, opiniones y expresiones: a) libertad de no revelar los pensamientos propios; b) protección contra alteraciones inaceptables del pensamiento y c) la obligación positiva del Estado de crear un ambiente habilitador para la libertad de pensamiento.[34]

El segundo elemento sobre la protección contra alteraciones inaceptables del pensamiento ha servido para determinar su relación con la autonomía cognitiva, es decir, dicha protección se refiere a la prohibición de la interferencia con la autonomía mental, la cual asume ausencia de coacción o modificación a través de alteraciones y manipulaciones indebidas. Esto no significa que garantice protección contra pensamientos de otros o de procesos cotidianos de persuasión, pero sí permite proponer posibles criterios para identificar forma indebidas de manipulación que deberían ser analizadas caso por casos, entre ellas, a) el consentimiento, b) la ocultación o falta de claridad sobre la influencia ejercida por contenidos, c) la asimetría de poder entre influenciador y titular de derechos, y d) el daño como manipulación inaceptable, distinguida de la influencia aceptable, en función de si existe intención de dañar o si tiene el efecto de infligir daño.[35]

El marco de atributos de la libertad de pensamiento propuesta por el Relator resulta útil para enmarcar la protección de la libertad de opinión y expresión frente a las condiciones necesarias en la vida democrática ligadas a la autonomía de las personas, como lo es contar con un ecosistema informativo libre, sin interferencias indebidas que puedan manipular el debate público y la deliberación, en donde sea posible expresarse, cuestionar y probar ideas y opiniones de manera dinámica. Por lo tanto, **consideramos fundamental que la RELE analice la forma en que los sistemas de IA alteran, afectan, influencian indebidamente e incluso manipulan el ecosistema informativo, afectando las libertades de opinión, expresión y pensamiento**.

El fenómeno de la "descarga cognitiva", por ejemplo, muestra cómo de manera creciente las personas abandonan tareas cognitivas fundamentales para desarrollar la capacidad de pensamiento crítico y de la memoria, que van desde la búsqueda de información, su análisis y verificación, la resolución de problemas lógicos, la redacción de ideas, entre otras, para delegarlas

---

33. Ibid, párr. 21.

34. Office for Democratic Institutions and Human Rights (ODIHR). *Think Again: Freedom of Thought in the Age of Artificial Intelligence*, OSCE, p. 12, 2025; Asamblea General de la Organización de las Naciones Unidas. "Informe provisional del Relator Especial sobre la libertad de religión o de creencias, Ahmed Shaheed", 5 de octubre de 2021, A/76/380.

35. Asamblea General de la Organización de las Naciones Unidas. "Informe provisional del Relator Especial sobre la libertad de religión o de creencias, Ahmed Shaheed", párras. 35-36, 5 de octubre de 2021,

sistemáticamente a los sistemas de IA. Al hacerlo, se deteriora la base práctica necesaria para el pensamiento crítico y la reflexión.[36] El fenómeno del "sesgo de automatización" se da cuando las dinámicas de deferencia a los sistemas de IA se naturalizan de manera que la agencia de las personas "se transfiere" a ellas. La actuación independiente, el razonamiento activo y la verificación crítica a las sugerencias de los sistemas automatizados queda desplazada, generando que en los procesos de decisiones asistidas por IA se tenga una deferencia automática, pero también que las opiniones y las creencias de las personas sean deferentes a las sugerencias o decisiones tomadas por la IA. Lo que se pone en juego en términos de agencia, autonomía y libertad de expresión es el propio proceso de pensamiento, al relegarlo a la IA en aras de la rapidez que proporciona para realizar distintas tareas automatizadas.[37]

---

36. Hao-Ping Lee et al., *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*, in PROCEEDINGS OF THE 2025 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 1 (April 26, 2025), https://dl.acm.org/doi/10.1145/3706598.3713778; E. F. Risko, S. J. Gilbert, Cognitive Offloading. Trends in Cognitive Sciences 20, 676–688 (2016): https://doi.org/10.1016/j.tics.2016.07.002; M. Gerlich, AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies (Basel, Switzerland) 15, 6 (2025): https://doi.org/10.3390/soc15010006; N. Kosmyna, E. Hauptmann, Y. T. Yuan, J. Situ, X.-H. Liao, A. V. Beresnitzky, I. Braunstein, P. Maes, Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task, arXiv [cs.AI] (2025): http://arxiv.org/abs/2506.08872; B. N. Macnamara, I. Berber, M. C. Çavuşoğlu, E. A. Krupinski, N. Nallapareddy, N. E. Nelson, P. J. Smith, A. L. Wilson-Delfosse, S. Ray, Does Using Atomartificial Intelligence Assistance Accelerate Skill Decay and Hinder Skill Development without Performers' Awareness? Cognitive Research: Principles and Implications 9, 46 (2024): https://doi.org/10.1186/s41235-024-00572-8; C. Zhai, S. Wibowo, L. D. Li, The Effects of over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review. Smart Learning Environments 11, 28 (2024): https://doi.org/10.1186/s40561-024-00316-7

37. L. J. Skitka, K. Mosier, M. D. Burdick, Accountability and Automation Bias. International Journal of Human-Computer Studies 52, 701–717 (2000): https://doi.org/10.1006/ijhc.1999.0349; K. Goddard, A. Roudsari, J. C. Wyatt, Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators. Journal of the American Medical Informatics Association: JAMIA 19, 121–127 (2012): https://doi.org/10.1136/amiajnl-2011-000089; T. Dratsch, X. Chen, M. Rezazade Mehrizi, R. Kloeckner, A. Mähringer-Kunz, M. Püsken, B. Baeßler, S. Sauer, D. Maintz, D. Pinto Dos Santos, Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology 307, e222176 (2023): https://doi.org/10.1148/radiol.222176; I. A. Qazi, A. Ali, A. U. Khawaja, M. J. Akhtar, A. Z. Sheikh, M. H. Alizai, Automation Bias in Large Language Model Assisted Diagnostic Reasoning among AI-Trained Physicians, medRxiv (2025); https://doi.org/10.1101/2025.08.23.25334280; F. Kücking, U. Hübner, M. Przysucha, N. Hannemann, J.-O. Kutza, M. Moelleken, C. ErfurtBerge, J. Dissemond, B. Babitsch, D. Busch, Automation Bias in AI-Decision Support: Results from an Empirical Study. Studies in Health Technology and Informatics 317, 298–304 (2024): https://doi.org/10.3233/SHTI240871; J. W. Ohde, L. M. Rost, J. D. Overgaard, The Burden of Reviewing LLM-Generated Content. NEJM AI 2 (2025): https://doi.org/10.1056/aip2400979; D. Lyell, E. Coiera, Automation Bias and Verification Complexity: A Systematic Review. Journal of the American Medical Informatics Association: JAMIA 24, 423–431 (2017): https://doi.org/10.1093/jamia/ocw105; R. Parasuraman, D. H. Manzey, Complacency and Bias in Human Use of Automation: An Attentional Integration. Human Factors 52, 381–410 (2010): https://doi.org/10.1177/0018720810376055; S. Passi, M. Vorvoreanu, "Overreliance on AI: Literature Review" (Microsoft, 2022): https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/; Z. Buçinca, M. B. Malaya, K. Z. Gajos, To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. Proceedings of the ACM on Human-Computer Interaction 5, 1–21 (2021): https://doi.org/10.1145/3449287;

3. El uso de la IA para dañar instituciones básicas de la democracia

Las democracias modernas necesitan de instituciones básicas para ejercer el poder estatal. Estas "instituciones críticas" son puentes entre el Estado y las personas, deben tratarlas como seres libres e iguales y cumplen un rol fundamental para garantizar bienes como la educación, la información y la comunicación, la paz, o la salud (universidades, prensa, instituciones médicas, instituciones de seguridad pública). Como el poder estatal se materializa a través de ellas, la democracia exige que éste *se justifique* y *se explique* a través de un lenguaje de "razones de interés público" relacionadas con los derechos de las personas y otros principios democráticos, los cuáles limitan la discrecionalidad y evitan la arbitrariedad del ejercicio del poder estatal, dando la posibilidad de cuestionar las decisiones estatales de manera transparente y abierta, y de rebatirlas en un marco amplio de deliberación. Todo esto queda en entredicho cuando se utilizan sistemas de IA en el ámbito público.[38]

Al implementar sistemas de IA que son opacos y sobre los que no existe transparencia significativa (no sólo general sino algorítmica), se elimina la posibilidad de revisar, escrutar, cuestionar y comprender la forma en que el poder público se ejerce (cuando estos sistemas se usan). Sin transparencia ni rendición de cuentas, cualquier noción de Estado de Derecho para la actuación de las instituciones se diluye, y la posibilidad de legitimar sus decisiones se anula porque pierden su justificación y explicación, corrompiendo los principios básicos del ejercicio democrático del poder estatal al perder sus bases procedimentales.[39]

De manera bastante abierta, la comunidad de grandes empresas de tecnología de IA ha dicho claramente que uno de sus objetivos es sustituir, precisamente, estos sistemas y procesos de gobernanza estatal, así como la propia estructura democrática por el modelo de los sistemas de inteligencia artificial para *agilizar* los procesos de toma de decisiones y prescindir prácticamente en todo del elemento humano, participativo y deliberativo que encuentran en ellos.[40]

En estos casos, no sólo se trata de que el diseño de estos sistemas va en sentido contrario de los principios fundamentales de una institución, como sucede en el caso de la prensa, en donde

---

38. Rawls, John. *Political Liberalism*. Columbia University Press, 1993 (revisited edition 2005); ver en particular "The Idea of Public Reason Revisited"; Hartzog, Woodrow & Silbey Jessica. *How AI Destroys Institutions*, 77 UC Law Journal (2026), p.3 Available at: https://scholarship.law.bu.edu/faculty_scholarship/4179

39. Nathalie A. Smuha. *Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law* (2024); Robert D. Putnam. *Bowling Alone: The Collapse and Revival of American Community*, 27 (2020); Julie E. Cohen, *Public Utility for What? Governing AI Datastructures*, 27 YALE J. L. & TECH. (2025); Kim Lane Scheppele, *The Life of the Rule of Law*, 20 ANN. REV. L. & SOC. SCI. 17, 20 (2024); Paul Gowder. *The Rule of Law in the Real World*, Cambridge University Press, pp. 12-20 (2016); Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 613-14 (2020).

40. Jill Lepore, *How We the People Lost Control of Our Lives, and How We Can Get It Back*, New York Times (Sept. 17, 2025), Disponible en: https://www.nytimes.com/2025/09/17/opinion/altman-ai-constiutional-convention.html

el compromiso con la veracidad y la búsqueda de la verdad son desplazados por la producción ilimitada de respuestas, la generación de información rápida y la complacencia de las personas que utilizan la tecnología (sin contar los usos malintencionados y dañinos que se le puede dar de forma deliberada para desinformar, manipular y engañar a la gente). Se trata de que tareas centrales que permiten la justificación, explicación y comunicación de las decisiones estatales son "descargadas" en sistemas que son opacos, incomprensibles, blindados del escrutinio público e irresponsables frente a la rendición de cuentas que las instituciones deben tener (o que esconden y desvían esa responsabilidad y *trasladan* esas decisiones a los desarrolladores que crean esas tecnologías y despliegan sus visiones del mundo en las mismas).[41]

Por un lado, la oscuridad en las empresas y tecnologías de IA -en clara tensión con los principios de transparencia y rendición de cuentas- es establecida con el ocultamiento que se da detrás de la aparente confianza declarativa de la producción de información (texto, imágen o video) que estos sistemas producen, mientras que esconden los juicios normativos establecidos por quienes diseñan dichas tecnologías y luego quedan por fuera de cualquier escrutinio público y rendición de cuentas, justificándose en el argumento utilitarista de la eficiencia y la velocidad.[42] Estos sistemas son incompatibles con la explicabilidad y predictibilidad necesarias de cualquier institución y medida que materializa el poder estatal y, más bien, establecen la incertidumbre como regla general de todos aquellos procesos en los que se involucran y son parte de un proceso más amplio de toma de decisiones.[43]

La IA erosiona los procesos de toma de decisiones que las instituciones deben tener para justificar y explicar las decisiones que toman al delegar tareas correspondientes a esos procesos a sistemas de IA, pasando esas etapas a las decisiones de los desarrolladores que elaboran esos sistemas. El análisis normativo, la reflexión sobre el contexto y las circunstancias particulares de cada caso, y el intercambio deliberativo necesarios para que el deber de rendición de cuentas sea cumplido, son sustituidos por los procesos inescrutables de la IA. Las decisiones sobre el acceso y el ejercicio de los derechos de las personas deben tomar en cuenta consideraciones mo-

---

41. Hartzog, Woodrow & Silbey Jessica. *How AI Destroys Institutions*, 77 UC Law Journal (2026), pp. 4-13. Available at: https://scholarship.law.bu.edu/faculty_scholarship/4179

42. Gerben Wierda, *Generative AI 'Reasoning Models' Don't Reason, Even If It Seems They Do*, R&A IT STRATEGY & ARCHITECTURE (June 8, 2025). Disponible en: https://ea.rna.nl/2025/02/28/generative-ai-reasoning-models-dont-reason-even-if-it-seems-they-do/; Julie Cohen, Oligarchy, State, and Cryptopia, 94 Fordham L. Rev. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5171050; Hartzog, Woodrow & Silbey Jessica. *How AI Destroys Institutions*, 77 UC Law Journal (2026), pp. 14-15

43. Danielle Keats Citron, Technological Due Process, 85 Wash. U. L. Rev. 1249 (2008); Frank Pasquale. *New Laws of Robotics: Defending Human Expertise in the age of AI*, Cambridge, Harvard University Press, pp. 119-144, 2020; John Kay & Mervyn King. *Radical Uncertainty: Decision-Making beyond the numbers*, Palgrave Macmillan; National Association for Business Economics, vol. 56(2), 2021.

rales, políticas y sociales cuya complejidad y matices no pueden ser subsumidos matemática, estadística y probabilísticamente por la idea de la "eficiencia neutral" de las tecnologías de IA.[44]

El carácter reflexivo de estas funciones requiere incorporar el contexto, las circunstancias particulares de cada caso y la complejidad de la realidad social, así como otros valores epistémicos como la empatía, la creatividad y el compromiso con el trato digno de las personas. Todo esto queda por fuera de los procesos de las tecnologías de IA, por lo que se pierde la valoración y el juicio humanos necesarios para la labor institucional del ámbito público.[45] La complejidad social y la especificidad particular que las instituciones deben atender, en su imbricación cualitativa, no puede ser estrechada y filtrada por la visión estadística de la IA.

> "La justicia se administra no en promedios sino en casos individuales [...]. Las narrativas son los medios por los que los seres humanos -como jueces, jurados o personas conduciendo sus asuntos cotidianos de la vida- ordenamos nuestros pensamientos y hacemos sentido de la evidencia que nos es dada. El estilo legal de razonamiento, esencialmente de inferencia lógica (abductivo), involucra la búsqueda por la "mejor explicación -por la narrativa más persuasiva de eventos relevantes al caso".[46]

V. El impacto de los sistemas de IA en la moderación de contenidos y el ejercicio amplio de la libertad de expresión

En línea con las solicitudes de transparencia que R3D ha hecho a a las empresas que moderan y curan contenidos en plataformas digitales, afirmamos que no existen salvaguardas para evitar discriminación o arbitrariedad en sus sistemas de moderación de contenidos. Esto aplica no sólo para las cuestiones relacionadas con el uso de IA en dichos sistemas sino en general para las distintas decisiones que toman al respecto, por ejemplo cuando deciden bloquear cuentas o eliminar contenidos por solicitudes de las autoridades. En distintos momentos, R3D intentó colaborar con Meta en investigaciones relacionadas con eliminación de contenidos en sus plataformas y con bloqueo de personas usuarias en las mismas, pero sistemáticamente la empresa se rehusó a entregar la información suficiente para poder de-

---

44. John Kay & Mervyn King. *Radical Uncertainty: Decision-Making beyond the numbers*, Palgrave Macmillan; National Association for Business Economics, vol. 56(2), 2021.

45. Michelle M. Mello & Sherri Rose, *Denial—Artificial Intelligence Tools and Health Insurance Coverage Decisions*, 5 JAMA HEALTH FORUM (2024). Disponible en: https://jamanetwork.com/journals/jama-health-forum/fullarticle/2816204; John W. Meyer, *The Effects of Education as an Institution*, 83 AM. J. SOCIO. 55 (1977); Dietrich Rueschemeyer, *Professional Autonomy and the Social Control of Expertise*, in SOCIOLOGY OF THE PROFESSIONS: LAWYERS, DOCTORS AND OTHERS 38 (Robert Dingwall & Phillip Lewis eds., Quid Pro Books 2014) (1983);

46. JOHN KAY & MERVYN A. KING, RADICAL UNCERTAINTY: DECISION-MAKING BEYOND THE NUMBERS, pp. 2010-11 (2021).

terminar posibles violaciones al derecho a la libertad de expresión, y meramente refirió a la información que no está desagregada ni es específica para poder determinar estos casos (que está en sus informes periódicos).

VI. Impactos de la IA en contextos migratorios y fronterizos

Diversos documentos publicados por el colectivo Guacamaya revelan información sobre el Sistema de Información para la Toma de Decisiones (SI2), desarrollado por el Estado Mayor Conjunto de la Defensa Nacional (EMCDN).[47] Este sistema centraliza información recopilada mediante labores de inteligencia y contiene perfiles de personas consideradas de interés para las Fuerzas Armadas: activistas sociales, comunicólogos, empresarios, políticos, miembros de caravanas migrantes, supuestos integrantes del crimen organizado y grupos terroristas. Los documentos mencionan planes para incorporar sistemas de IA al SI2 en el futuro.

Hasta la fecha, no existe evidencia documental de que las Fuerzas Armadas hayan evaluado el impacto de estos sistemas sobre derechos fundamentales como el debido proceso, la libertad de expresión, de asociación y de reunión, ni de que hayan adoptado medidas para prevenir sus violaciones. Lo que sí es claro es que, ante investigaciones por parte de medios de comunicación, al ser cuestionados por vía de Solicitudes de Acceso a la Información, distintas autoridades militares, tal como la Guardia Nacional, así como autoridades de seguridad pública, tal como la Fiscalía General de la Nación, reconocieron utilizar sistemas de inteligencia artificial para realizar sus tareas sin revelar cuáles, en qué consistían ni con qué fines específicamente, al clasificar dicha información como reservada por motivos de seguridad nacional, impidiendo así el escrutinio público de la misma.[48]

VII. Los centros de datos para la IA y sus impactos medioambientales

El gobierno federal y los gobiernos locales en México han permitido el establecimiento de al menos doce centros de datos para la inteligencia artificial. El gobierno de Querétaro no sólo ha dado facilidades económicas e impositivas a grandes empresas tecnológicas extranjeras para

---

47. Ver: https://r3d.mx/wp-content/uploads/Documentacion-Proyecto-COC-EMCDNv2.1.pages, https://r3d.mx/wp-content/uploads/Anexo-D-Presentacion-Sistema-Informacion-07-Oct-2022-SI2.pptx y https://r3d.mx/wp-content/uploads/Anexo-D-Lista-concentrada-de-Historias-de-Usuario.xlsx

48. Estrada, Andrés."El gobierno mexicano ya usa la IA, pero no hay normas ni estrategia", *WIRED*, 3 de marzo de 2026. Disponible en: https://es.wired.com/articulos/el-gobierno-mexicano-ya-usa-la-ia-pero-no-hay-normas-ni-estrategia

invertir en este tipo de infraestructura, sino que ha facilitado que estas eludan requisitos legales relacionados con los impactos medioambientales que esta infraestructura tiene, tal como el requisito de presentar una manifestación de impacto ambiental y el de pagar impuestos medioambientales correspondientes.[49]

La manera en que las autoridades han arreglado este desvío regulatorio es a partir de una interpretación que ignora la dimensión medioambiental de los centros de datos, que consiste en tomar estas infraestructuras como infraestructura relacionada con la prestación de servicios digitales, en vez de considerarlas como empresas industriales que son de carácter (o podrían serlo) contaminante, por lo que deben someterse a las obligaciones y regulaciones de este ámbito.

El titular de la Secretaría de Desarrollo Sustentable (SDS) de Querétaro, Marco del Prete, reconoció la exención de presentar las manifestaciones de impacto ambiental porque estas empresas no podían ser consideradas como industrias sino como empresas prestadoras de servicios: "El data center es una empresa de servicios que no procesa insumos ni genera emisiones directas. No tiene por qué obtener una manifestación de impacto". Existen registros que datan del año 2022, en los que este organismo (SDS) reconoció explícitamente que el centro de datos operado por la empresa Microsoft no requería un informe de impacto ambiental ni cumplir con obligaciones de transparencia sobre impactos medioambientales porque sus actividades "no se consideran fuentes fijas de emisiones".[50]

En la práctica, las permisiones e irresponsabilidad de las empresas (no sólo *Microsoft*, sino también *Google* y *Amazon*, dado que también han invertido en este tipo de infraestructuras en parques industriales en el Estado de Querétaro) relacionadas con los centros de datos violan tanto disposiciones locales como el Código Ambiental, que obliga a este tipo de sujetos a realizar estudios de impacto ambiental para poder informar a las comunidades y evaluar posibles obligaciones en este ámbito, como también obligaciones internacionales relacionadas con la información medioambiental para las comunidades afectadas, tales como aquellas contenidas en el Acuerdo de Escazú. Además de las empresas actualmente en Querétaro, la propia presidenta Claudia Sheinbaum explicó en septiembre del año pasado que la empresa *CloudHQ* invertiría casi 5 mil millones de dólares para construir seis centros de datos más en este Estado.[51]

---

49. Tovar, Miguel. "Centros de datos atraídos a México pueden evadir requisitos medioambientales", WIRED, 14 de octubre de 2025. Disponible en: https://es.wired.com/articulos/centros-de-datos-atraidos-a-mexico-pueden-evadir-requisitos-medioambientales

50. Ídem.

51. Ídem.

La gravedad institucional de la ausencia de regulación y de la irresponsabilidad de estas empresas frente al medio ambiente y las comunidades afectadas por este tipo de infraestructuras en México se corresponde con un modus operandi de estos sujetos frente a las comunidades vulnerables, en donde la marginalización es la regla y la necesidad descomunal de poder computacional de los centros de datos hace inevitable la destrucción ambiental.[52] En el caso de Querétaro, la región está sujeta a un estrés hídrico extremo y el acceso al agua no es una realidad para muchas comunidades. Sin embargo, las decisiones de los gobiernos involucrados han puesto por encima los intereses económicos de la inversión en centros de datos sobre los derechos de las personas en este Estado.

VIII. El uso de la IA para ataques de ciberseguridad y robo de datos

De manera creciente, las tecnologías de inteligencia artificial se utilizan para diseñar ataques maliciosos a infraestructuras críticas de los gobiernos, poniendo en riesgo bienes como los datos las personas, incluidos los datos sensibles o datos que pueden ser utilizados con motivos políticos para afectar derechos humanos.

Un caso reciente en México para ilustrar el riesgo del uso de estas tecnologías para los derechos de las personas es el del hacker que utilizó un *chatbot* de *Anthropic PBC* para robar información confidencial y personal en manos de autoridades del gobierno mexicano. El ataque fue dirigido principalmente a la autoridad fiscal, el Servicio de Administración Tributaria (SAT), y a la autoridad electoral federal, el Instituto Nacional Electoral (INE). El atacante robó más de 150 Gigabytes de información fiscal y de votantes, además de atacar también a distintas autoridades locales en estos y otros ámbitos. Este ataque se realizó evitando las supuestas barreras sobre usos indebidos de *Claude*, y logró evitar los filtros para que el sistema de IA realizara las tareas ilegales que permitieron acceder a los datos protegidos. Según la investigación sobre este caso, cuando el *hacker* encontraba limitaciones en *Claude*, utilizaba *Chat GPT* de *OpenAI* para ampliar los datos y burlar la limitación.[53]

---

52. Adam Zewe, *Explained: Generative AI's Environmental Impact*, MASS. INST. TECH. NEWS (Jan. 17, 2025), https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117; Shaolei Ren & Adam Wierman, *The Uneven Distribution of AI's Environmental Impacts*, HARV. BUS. REV. (July 15, 2024), https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

53. Bloomberg. "Hacker Used Anthropic's Claude to Steal Mexican Data Trove", February 25, 2026. Ver en: https://www.bloomberg.com/news/articles/2026-02-25/hacker-used-anthropic-s-claude-to-steal-sensitive-mexican-data?embedded-checkout=true