Lima, 31 de Mayo del 2024

Comentarios al Proyecto de Reglamento de la Ley 31814, Ley que promueve el uso de la IA en favor del desarrollo económico y social del país - Perú

Los siguientes comentarios surgen del trabajo conjunto entre las organizaciones de la sociedad civil Access Now e Hiperderecho.

Access Now defiende y extiende los derechos digitales de las personas y comunidades en riesgo alrededor del mundo. Access Now trabaja en asociación con actores locales para promover la agenda de los derechos humanos en el uso, desarrollo y gobernanza de las tecnologías digitales, interviniendo en los casos donde éstas impactan negativamente.

Hiperderecho es una asociación civil peruana sin fines de lucro dedicada a promover e investigar el respeto de los derechos humanos en entornos digitales, conformada por abogadas y especialistas en tecnología. Como parte de nuestro trabajo, estudiamos todas las iniciativas de política pública que puedan impactar el ejercicio de derechos y libertades en estos ámbitos.

I. Comentarios generales

1. Consideraciones sobre el alcance de la norma hacia las empresas

De manera introductoria, se deben señalar cuatro comentarios transversales al Proyecto normativo. En primer lugar, respecto al alcance de la norma, en el Proyecto se señala, en el artículo 2, el ámbito de aplicación entre los cuales están las entidades de la administración pública, las empresas que realizan actividad empresarial del Estado, las empresas públicas y, en el último inciso “las organizaciones de la sociedad civil, ciudadanos, academia y el sector privado”. Si bien de la disposición se interpreta que el sector privado incluye a las empresas privadas, es necesario que este alcance figure de forma clara, sin lugar a interpretaciones que puedan mermar el alcance del proyecto normativo. Inclusive, se debe destacar la alusión en el citado inciso al artículo 6 del Decreto de Urgencia que crea el Sistema Nacional de Transformación Digital[1]. El cual afirma: “Los principios, normas y procedimientos que rigen la materia de Transformación Digital son aplicables a [...] las organizaciones de la sociedad civil, ciudadanos, empresas y academia en lo que corresponda”.

No está de más señalar que excluir del alcance del Proyecto normativo a las empresas, sería incurrir en un tratamiento diferenciado injustificado entre las empresas privadas que desarrollan e implementan IA y las empresas públicas o los órganos de la Administración Pública ¿por qué las primeras deberían estar exentas de estas obligaciones?

[1] Presidente de la República (2020) Decreto de Urgencia N°006-2020.

Ello, sobre todo siguiendo el enfoque de riesgos: los sistemas de IA tienen la capacidad de generar impactos significativos negativos sobre los derechos humanos, independientemente de que estos sean implementados por el sector empresarial privado o estatal. En otras palabras, si el criterio central del Proyecto es la clasificación del tipo de riesgo, esta se hará para todos los sistemas de IA, siendo indiferente, para esta clasificación y las consecuentes obligaciones, el sector en el que se desarrolla o implementa. Lo contrario sería admitir en el sistema jurídico un trato diferenciado injustificado hacia las empresas privadas, lo que es incompatible con el derecho-principio de igualdad y prohibición de la discriminación.

2. No se han previsto plazos para la adaptación de sistemas de IA ya existentes

En segundo lugar, el Proyecto no prevé plazos para la adaptación de sistemas de IA ya existentes. Naturalmente, la implementación de las obligaciones y sistemas de supervisión que establece el Proyecto deben ir acompañados de un cronograma de plazos a fin de que las disposiciones puedan ser implementadas en un tiempo razonable. De hecho, el Proyecto menciona que los instrumentos para el uso y desarrollo de IA que dispone serán desarrollados de forma gradual y contínua (artículo 9.2). Sin embargo, esta previsión se hace respecto a disposiciones específicas y de forma enunciativa.

Otras normativas del sector sí establecen plazos claros - e identifican objetivos y actores responsables- para cambios significativos asociados a la transformación digital. Por mencionar dos ejemplos, el Texto Único Ordenado de la Ley 27806, Ley de Transparencia y Acceso a la Información Pública[2], dispone de plazos para la implementación de los portales de internet de las entidades públicas, dependiendo del tipo de entidad. Por otro lado, la Resolución Administrativa que reglamenta el Expediente Judicial Electrónico (EJE)[3], establece etapas para su implementación, una primera etapa piloto y una segunda de implementación progresiva de acuerdo a la jurisdicción y especialidad de cada órgano.

Inclusive la Ley de IA de la Unión Europea[4], que ha sido referente para este Proyecto normativo, establece también plazos, y diferenciados, en función de las distintas acciones y obligaciones. Por ejemplo, establece un plazo de seis meses para los tipos de riesgo inaceptable, y un plazo de dos años para que entren en vigor las obligaciones sobre los sistemas de IA de alto riesgo.

3. No hay regulación asimétrica o diferenciada entre tipos y tamaños de empresa

Un tercer comentario transversal que se debe apuntar sobre el Proyecto es el referido a los tipos y tamaños de empresas. De una lectura global, se puede observar que el Proyecto está pensado para las grandes empresas e industrias que desarrollen e implementen sistemas de

[2] Ministerio de Justicia y Derechos Humanos (2019). Decreto Supremo 021-2019-JUS.
[3] Consejo Ejecutivo del Poder Judicial (2017). Resolución Administrativa 228-2017-CE-PJ.
[4] European Parliament (2024). Artificial Intelligence Act. P9_TA(2024)0138.

IA. Esto es posible evidenciarlo en la figura del implementador y sus obligaciones (artículos 19 y 20). Entre las cuales se encuentra contar con un equipo diverso y multidisciplinario para el diseño, desarrollo e implementación del sistema basado en IA, e identificar y minimizar los sesgos de los algoritmos o bases de datos. Inclusive, en el artículo 20 se presupone que el implementador es un equipo dentro del cual existen los rangos de alta dirección y equipo de operación.
De esta forma, es recomendable incluir, ya sea como un principio, como una sección independiente o en diversos artículos esta regulación diferenciada entre tipos y tamaños de empresa. A fin de que la ejecución normativa no genere un impacto desproporcionado y se pueda preservar el equilibrio entre las obligaciones y los sujetos obligados.

Por otro lado, resulta prudente señalar en este punto que el principal referente de este Proyecto, la Ley de IA de la Unión Europea, establece diferencias en las obligaciones, especialmente para las pequeñas y medianas empresas (pymes) y para las empresas emergentes. De hecho, en su artículo 62 señala las medidas dirigidas a las pymes y empresas emergentes, entre las cuales se encuentra el acceso prioritario a los espacios controlados de pruebas para la IA, canales de comunicación específicos, y tasas diferenciadas para la evaluación de conformidad -de haber cumplido con los requisitos de acuerdo al tipo de riesgo-, en función al tamaño de la empresa, tamaño del mercado, entre otros indicadores.

4. **Resulta necesaria la inclusión de la obligación de llevar adelante estudios de impacto a Derechos Humanos**

Si bien existen consideraciones en torno al impacto que ciertos sistemas pueden tener sobre determinadas poblaciones, no se hace mención en ningún momento a la necesidad de llevar adelante estudios de impacto a derechos humanos como requisito previo al despliegue que ahonden sobre estas potenciales consecuencias con metodologías propicias para tal fin. Los estudios de impacto son muy necesarios para identificar, por un lado, los riesgos intrínsecos vinculados al tipo de sistema que se pretende poner en uso, como así también para ofrecer la evidencia que eventualmente permita una clasificación adecuada del sistema dentro del esquema propuesto. La reglamentación propuesta carece de un mecanismo básico, presente en la Ley de IA de la Unión Europea (artículo 27), de la cual la reglamentación en análisis es al menos un reflejo. Más consideraciones en torno a estrategias y estándares para conducir estas evaluaciones de impacto pueden encontrarse en el documento de Access Now titulado "Human Rights Impact Assessment for AI: Analysis and Recommendations"[5].

5. **La importancia de precisar el abordaje a la IA desde la ética**

[5] Brandie Nonnecke y Philip Dawson (2022). Human Rights Impact Assessments for AI: Analysis and Recommendations. Access Now. Disponible en https://www.accessnow.org/wp-content/uploads/2022/11/Access-Now-Version-Human-Rights-Implications-of-Algorithmic-Impact-Assessments_-Priority-Recommendations-to-Guide-Effective-Development-and-Use.pdf

Se recomienda un uso más acotado y preciso del concepto de lineamientos éticos y de la ética en general (presentes en los arts.1, 4.d, 4.j, 8.d, 10.1, 12.2.f, Capítulo IV,. 20.1, 20.3, 23 y 32). El abordaje desde la ética sobre el desafío regulatorio y técnica de la IA tiene extensa difusión y ha sido ampliamente aceptado y desarrollado. Sin embargo, es necesario comprender cómo y cuándo es útil servirse de la ética y cuando esta se convierte en una etiqueta carente de sentido y efectos.

En primer lugar, la ética es una disciplina de la filosofía. Esto supone que su abordaje debe partir de la duda o desde el escepticismo, y que no necesariamente se debe arribar a conclusiones ni mucho menos a certezas. De esta manera, la pregunta por la ética de la IA es una que no debe ni puede justificar acciones administrativas ni corporativas, debiendo estas siempre basarse en el derecho positivo, en los principios fundamentales y en el respeto a los derechos humanos. Ejemplo de esto es como la ética de la IA presenta diferentes ángulos de aproximación y que existen más de 170 iniciativas de propuestas éticas para la IA que pueden consultarse en el Inventario administrado por la organización de la sociedad civil Algorithm Watch[6].

Por otro lado, si bien la ética puede servir como guía y como proceso de reflexión, esta dependerá del contexto y de los intereses en pugna. Ejemplo de esto es como la UNESCO, organismo que viene desarrollando estándares para la IA desde la ética, propone un mecanismo de análisis y diagnóstico para identificar los desafíos específicos que existen en un país determinado[7].

La mención de la “ética” no debe perder sentido debido a su repetición o a partir de la ambigüedad de su mención. Como por ejemplo, en el artículo 20.3, donde se señalan a los “estudios de impacto ético”, que no tienen un desarrollo específico que permita comprender sus objetivos o alcances. Razón por la cual debemos sospechar de ellos y señalar a los estudios de impacto a derechos humanos como el mecanismo adecuado para evaluar riesgos y medidas de mitigación adecuadas frente a un sistema. Por este motivo y para que el abordaje de la ética sea una herramienta efectiva, se recomienda dosificar su señalamiento.

Además, es necesario que la ética no reemplace ni confunda la jerarquía que debe sostenerse en el reconocimiento y observación de los derechos fundamentales, y de los estándares internacionales de derechos humano, los cuales son y deben seguir siendo la referencia por excelencia a tener presente durante todo el ciclo de vida de los sistemas de IA, y de la cual debe derivar todo diseño regulatorio.

---

[6] Algorithm Watch. AI Ethics Guidelines Global Inventory, disponible en https://inventory.algorithmwatch.org/

[7] UNESCO. (28 de agosto de 2023). Readiness Assessment Methodology: a tool of the Recommendation on the ethics of Artificial Intelligence. Disponible en https://www.unesco.org/en/articles/readiness-assessment-methodology-tool-recommendation-ethics-artificialintelligence

II. Comentarios específicos

1. El implementador no está correctamente identificado y no se distinguen los sujetos obligados

De acuerdo con las definiciones, el implementador es “toda persona natural o jurídica que utiliza un sistema basado en inteligencia artificial excepto cuando su uso se da exclusivamente para actividades personales.” Esta definición es especialmente problemática porque la figura del implementador es medular y por ende resulta cuestionable por dos razones. En primer lugar, porque a partir de ella no se puede diferenciar al implementador de la persona que usa la IA (persona usuaria), sobre todo a partir de la ausencia de conceptualización de lo que supone una “actividad personal” y de lo que es una “actividad que excede lo personal”. Podría interpretarse que la definición de implementador excluye a aquellas personas cuyo uso del sistema basado en IA se realice exclusivamente para “actividades personales”. De cualquier manera, este concepto resulta ambiguo, impreciso y amplio. Posiblemente lo que se intentaba hacer en el Proyecto era exceptuar de la aplicación de la figura del implementador a quienes usen sistemas de IA para fines exclusivamente relacionados con su vida privada o familiar. No obstante, eso no supera la discusión, porque de ser así, la narrativa estaría pensada para excluir a las personas usuarias. Pese a que al inicio se define al implementador como aquella persona “que utiliza” un sistema basado en IA.

Y, en segundo lugar, en todo el Proyecto se señala una serie de obligaciones para el implementador, como es el deber de tomar las medidas de gestión adecuadas al nivel de riesgo (arts. 14.1 y 16.1), demostrar el cumplimiento de sus obligaciones (art.16.2), incorporar medidas para la protección de la privacidad y los datos personales (art.18), entre otras que incluyen incluso su responsabilidad penal, civil, o administrativa derivada del uso y desarrollo de sistemas de IA.

Sumado a ello, en el Proyecto no se logra diferenciar a los distintos sujetos obligados, que deberían ser, además del implementador, el desarrollador, el distribuidor o el importador del sistema de IA. Por ejemplo, en el artículo 20.2 se establecen obligaciones entre las cuales figuran actividades asociadas no sólo a la implementación, sino al diseño y desarrollo de un sistema de IA. Asimismo, en otros artículos se utiliza como sinónimos proveedor e implementador (arts 26 y 31), o se mencionan al distribuidor e importador (arts. 16.2 y 21), aunque estas figuras no aparecen entre las definiciones.

Al respecto, resulta interesante ver que, comparativamente, la Ley de IA de la Unión Europea distingue en su aplicación (art. 2) a los proveedores, a los responsables del despliegue o implementadores (deployers), a los importadores y distribuidores, a los fabricantes de productos, entre otros. Tomando en cuenta cada uno de estos actores se establecen pautas de coordinación entre ellos, así como obligaciones diferenciadas. Asimismo, la figura de cada uno

se desarrolla en el artículo sucesivo (art.3) de definiciones, siendo la definición del implementador la misma que aparece en el Proyecto, lo que desencadena la pregunta de por qué sólo se definió esta y se centraron las obligaciones en esta, sin desarrollar las demás.

De esta forma, resulta imprescindible que en el Proyecto se pueda clarificar la figura del implementador, del importador y/o distribuidor, del fabricante/desarrollador, y que se identifique los distintos sujetos obligados respecto de las actividades y las obligaciones de cada uno.

2. Existe una contradicción en el ámbito de aplicación

En el art. 2 del Proyecto, dedicado al ámbito de aplicación, no se reconoce el sistema jurisdiccional. Sin embargo, en el art. 14.2.g, se menciona como un riesgo alto de sistemas de IA, "el apoyo en la toma de decisiones, interpretación normativa o de los hechos por una autoridad jurisdiccional".

Ciertamente en la Ley 31814, que promueve el uso de la IA en favor del desarrollo económico y social del país, se menciona que "es de interés nacional [...] el fomento del desarrollo y uso de la IA para la mejora de [...] la justicia"; y sobre esta base la Corte Superior de Justicia de Lima aprobó el 2023 la implementación del "Laboratorio de Inteligencia Artificial de la Corte Superior de Justicia de Lima" [8]. No obstante, este Laboratorio tiene objetivos muy puntuales: (i) el desarrollo de capacidades y conocimientos en el uso responsable y ético de la IA y (ii) la generación de soluciones innovadoras para los desafíos del servicio de administración de justicia en el distrito judicial de Lima. En otras palabras, el Laboratorio no incluye, ni entre sus objetivos ni entre sus atribuciones, actividades como el uso de IA para el apoyo en la toma de decisiones o para la interpretación normativa. Actividades que resultan complejas en términos de protección de derechos humanos, sobre todo los relacionados con la tutela jurisdiccional efectiva y el debido proceso. Así, de incorporarse el sistema jurisdiccional en el ámbito de aplicación del Proyecto, se podría estar produciendo una afectación en el principio de separación de poderes.

Además, si bien en el ámbito jurisdiccional están avanzando proyectos piloto e iniciativas de IA como el proyecto "Justo"[9], para la elaboración de autos, en supuestos limitados y sin pronunciamiento de fondo, esto acarreará la necesidad de su propia regulación. Especialmente tomando en cuenta que el universo judicial tiene prioridades y objetivos diferentes, de acuerdo a su naturaleza, y que por ello puede presentar otro tipo de riesgos para los derechos humanos. Por ende, se sugiere que el ámbito de aplicación del Proyecto se restrinja al ámbito administrativo.

---

[8] Presidencia de la Corte Superior de Justicia de Lima (2023). Resolución Administrativa 000620-2023-P-CSJL-PJ.

[9] Consejo Ejecutivo del Poder Judicial (2022). Resolución Administrativa 000273-2022-CE-PJ, que aprueba el despliegue piloto del proyecto "Justo" robot asistente judicial en el Módulo Integrado en Violencia Contra la Mujer e Integrantes del Grupo Familiar de la Corte Superior de Justicia de Lima Norte.

3. **Consideraciones respecto a los "principios rectores para el uso y desarrollo de la inteligencia artificial"**

Si bien es loable incorporar un artículo que ahonde en los principios que deben regir la interpretación y aplicación de la reglamentación, es necesario presentar algunas consideraciones.

En el **principio de rendición de cuentas y supervisión humana** (art. 4.k) se establece: "Los sistemas basados en IA cuentan con mecanismos para la rendición de cuentas por su funcionamiento y uso [...] cuando las circunstancias lo ameriten, se incorporan mecanismos de supervisión humana a los sistemas basados en IA". Es decir, en el artículo se dispone que la supervisión humana es la excepción y no la regla. Sin embargo, a fin de evitar la ambigüedad es recomendable seguir el enfoque de gestión de riesgos y la coherencia con otros artículos. En esta misma línea, podríamos preguntarnos, por ejemplo, ¿cuándo las circunstancias ameritan contar con supervisión humana?, ¿para qué supuestos se aplica la excepción del mecanismo de supervisión humana?.

Así, se sugiere realizar una corrección para especificar en el art. 4.k que "[...] para los sistemas basados en IA de riesgo alto, se incorporan mecanismos de supervisión humana". Sería una modificación de forma principalmente, pues la supervisión humana ya se encuentra prevista como obligación para los sistemas basados en IA de riesgo alto, con el nombre de control humano, en el art. 16.1.h.

Sumado a lo anterior, sería adecuado esclarecer o profundizar en lo que consiste la supervisión humana, quién es la persona o grupo encargado de cumplir esta función, y cuál es el perfil o requisitos que se requieren para cumplir satisfactoriamente con la supervisión. También sería oportuno indicar si existirán mecanismos de control que permitan verificar el cumplimiento de esta obligación.

Respeto al **principio de transparencia y explicabilidad** (art. 4.i), el derecho reconocido a las personas usuarias por el cual pueden requerir una explicación sobre cómo funciona un determinado sistema y sobre cómo se ha alcanzado cierto resultado, es sin lugar a dudas una prerrogativa fundamental para acreditar la eventual responsabilidad civil y/o penal de los proveedores de estas tecnologías, como así también para asegurar la capacidad de supervisión humana sobre el funcionamiento de los mismos. De cualquier manera, es fundamental que la reglamentación genere los mecanismos necesarios para que dicho principio pueda ser llevado a cabo y no quede en una expresión de deseo. Debido a que la misma no depende de una norma que lo habilite o prescriba, sino de la introducción de mecanismos técnicos que deben ser considerados desde el diseño, los cuales hacen parte de un debate entre expertos que parecen no encontrar estándares para la persecución de tal fin. Por este motivo, llama la atención que la reglamentación propuesta asegure "la transparencia y

explicabilidad de los sistemas basados en inteligencia artificial, como la incorporación de medidas de transparencia algorítmica y de trazabilidad". Instamos a modificar la redacción de este principio de tal forma que se ahonde sobre el vínculo entre el derecho a recibir una explicación y la responsabilidad civil y/o penal que surge del daño causado. Es decir, a ir más allá del mero entendimiento para abrazar la naturaleza riesgosa de cualquier sistema al que se le delegan funciones o tareas, fortaleciendo el principio de responsabilidad que se enuncia en el numeral j del mismo artículo.

Por último, se sugiere agregar en la lista de principios al **principio de múltiples partes interesadas**, que se encuentra recogido en diversos artículos (art. 7.3, art. 9.2, art. 20.3, art. 20.5 y art. 32.3) y se reconoce como un enfoque en el art. 6.e. Inclusive, este se encuentra en la lista de principios para el desarrollo y uso de la IA, en el Título Preliminar de la Ley 31814 con el nombre de enfoque de pluralidad de participantes.

Cabe recordar además que este principio ya ha sido reconocido por el Poder Ejecutivo en otras normas del sector digital, como en el artículo 2.b del Decreto que crea el Sistema Nacional de Transformación Digital[10], con la denominación de compromiso y participación. El cual, en estricto, se refiere a que la "toma de decisiones, diseño de políticas y entrega de servicios digitales se realice utilizando enfoques, métodos o técnicas colaborativas" que atiendan las demandas y necesidades de múltiples sectores del ecosistema digital.

4. **Sería oportuno aclarar las acciones de supervisión de la Autoridad Nacional**

En el artículo 6.b se menciona entre las acciones de la Autoridad Nacional para el uso y desarrollo de la IA: "evaluar y supervisar la adopción de los sistemas basados en IA en el territorio nacional". Sin embargo, esta disposición podría resultar vaga o imprecisa. Por lo que consideramos que sería conveniente explicitar específicamente las acciones legales de supervisión que implementará la Secretaría de Gobierno y Transformación Digital, con el objetivo de fiscalizar el correcto cumplimiento de las distintas obligaciones expuestas en el Proyecto.

5. **El registro público como sistema de transparencia para monitorear los sistemas que usen o desarrollen IA**

El artículo 8.2.d indica que se deberá "Enviar trimestralmente la lista de los sistemas basados en inteligencia artificial que usen o desarrollen al canal digital que habilite la Secretaría de Gobierno y Transformación Digital para su difusión permanente en la sede digital de la Presidencia del Consejo de Ministros, conteniendo como mínimo, el tipo de sistema, la finalidad prevista y la persona responsable". Este artículo introduce una medida de transparencia que es oportuna, sobre todo si es desarrollada en mayor medida para incorporar otras consideraciones que la fortalezcan. En primer lugar, será oportuno regular la creación y administración de un

[10] Presidente de la República (2020) Decreto de Urgencia N°006-2020.

registro público que sea accesible de forma gratuita y que permita su lectura mecánica por parte de computadoras. Luego, se recomienda sumar a la información mínima requerida la siguiente:

- Información relativa al contexto en que un sistema es desplegado (es decir, las tareas específicas que le han sido delegadas),
- Documentación técnica actualizada incluyendo información sobre qué datos han sido utilizados para entrenar y qué datos requiere procesar el sistema para su funcionamiento.Las Medidas de seguridad implementadas, que deberán ser adecuadas y en sintonía con el estado de la técnica.
- Instrucciones de uso del sistema lo suficientemente claras para ser comprendidas por el público en general.

Con estos agregados, el artículo 8 en general supondría un avance regulatorio que permitirá a la ciudadanía acceder de mejor manera a la información relacionada a los sistemas desplegados, incluso pudiendo llevar adelante análisis computacionales que respalden la eventual toma de decisiones tanto desde el gobierno, como desde la sociedad en general.

6. La definición de IA generativa se confunde con otros términos

En el art. 26 del Proyecto, referido a las medidas de seguridad en relación con los sistemas basados en IA generativa, se define a la IA generativa como "sistemas basados en IA que cuentan con la capacidad de generar o alterar textos, imágenes, videos y audios, que puede inducir erróneamente a la creencia de que es auténtico o verídico". Esta definición resulta parcialmente correcta por dos razones.

Primero, porque si bien la IA generativa puede crear o generar contenidos como los descritos, no se limita a estos, podría incluirse también la generación o creación de códigos de software, por ejemplo. Segundo, porque la IA generativa suele definirse[11] como la creación o generación, pero no como la alteración o manipulación de contenidos audiovisuales con el fin de falsificar o engañar, lo que está referido más bien a los deepfakes y fake news, problemas que ciertamente se ven facilitados por la IA generativa, pero esta no se limita a ellos. Por lo que sugerimos se pueda aclarar esta definición.

7. Análisis sobre el enfoque basado en riesgos desde la perspectiva de los derechos humanos

El proyecto de reglamentación sigue el diseño de la recientemente aprobada ley de IA de la UE en cuanto a que diferencia los sistemas de IA a partir de una matriz que diferencia usos y

[11] UNESCO (2024). Guía para el uso de IA generativa en educación e investigación (p.8); UNESCO (2024). La violencia de género facilitada por la tecnología en la era de la IA generativa (p.10); y European Parliament (2024). Artificial Intelligence Act. P9_TA(2024)0138 (pp. 92 y 96).

finalidades específicos. Este diseño no está exento de críticas. En primer lugar, el enfoque adecuado para una regulación de uso de estas tecnologías será aquel basado o enfocado en derechos fundamentales. Cuando se diseñan políticas públicas a partir de enfoques basados en riesgos, se negocian las garantías a los derechos humanos a partir de la premisa de que estas deben equilibrarse con otros valores como los de la innovación. Es posible y además necesario fomentar una cultura de la innovación que tenga como centro no solo los derechos fundamentales de los individuos, sino también una búsqueda de vida en sociedad más armónica y sostenible. Incluso cuando se decide por diseños basados en riesgos, se suelen proponer ciertas excepciones o condiciones que justifican el despliegue de un sistema considerado de "alto riesgo" o incluso de "riesgo inaceptable", lo que supone habilitar el despliegue de sistemas riesgosos, por ejemplo:

- Por razones de seguridad nacional o control migratorio.
- Al mismo tiempo, otros sistemas que se reputan de riesgo alto o inaceptable quedan sujetos a prohibiciones o al cumplimiento de requisitos sólo cuando podrían ocasionar consecuencias negativas o tener un efecto perjudicial sobre los sujetos.
- Todos los diseños que imponen requisitos o prohibiciones que eventualmente pueden ser obviados, ya sea porque se configura un supuesto peligro para la seguridad nacional o porque potencialmente no generan efectos perjudiciales, **suponen guías o hojas de ruta para evitar el acatamiento a la normativa propuesta.**

Sugerimos abandonar el diseño basado en riesgos y explorar mecanismos que garanticen, más allá de las necesarias medidas de mitigación de riesgos que deben existir, un diseño reglamentario basado en los derechos humanos como premisa inclaudicable del objetivo detrás de legislar en torno a los sistemas de IA. De cualquier manera, nos detendremos a continuación en el articulado específico que presenta la particular forma en la que se propone la matriz de riesgos.

8. Críticas a la lista de sistemas considerados de riesgo alto

En el artículo 14.2 se describen los sistemas considerados de "riesgo alto", los cuales merecen una crítica y un abordaje desde la perspectiva de los derechos fundamentales.

8.1. Los sistemas de IA que asisten a la identificación biométrica y categorización de personas naturales

El numeral a) señala a los sistemas que asisten a la "identificación biométrica y categorización de personas naturales". Lo primero que tenemos que señalar es que no se han previsto definiciones específicas en la propuesta para conceptualizar a la "identificación biométrica" ni a la "categorización de personas naturales". Proponemos las siguientes definiciones:

- **Sistema de identificación biométrica:** sistema de inteligencia artificial que utiliza datos relacionados con características físicas, fisiológicas o conductuales de una persona

natural a los efectos de identificar a una persona a través de un proceso de comparación utilizando como referencia una base de datos preexistente.

- **Sistema de categorización biométrica:** sistema de inteligencia artificial que utiliza datos relacionados con características físicas, fisiológicas o conductuales de una persona natural a los efectos de asignarla a categorías específicas que puedan inferirse razonablemente de dichos datos.

En segundo lugar, criticamos la falta de precisión del numeral, que supone que todo sistema de categorización biométrica será de riesgo alto. Es preciso indicar que existen sistemas que presumen de su capacidad de categorización biométrica, asignado categorías como "preferencia política", "orientación sexual", o de "categorización basada en emociones" cuya efectividad es incomprobable (para no decir imposible) y que deben ser definitivamente consideradas de riesgo inaceptable y quedar totalmente prohibidas, ya que se encuentran en directa contradicción con los derechos humanos.

Nótese que sin haber ahondado en la cuestión de la identificación biométrica, es fácilmente deducible que lo mismo que se ha expuesto para la categorización biométrica sirve para la identificación, ya que existen contextos donde una potencial identificación se encuentra en directa contradicción con el principio de inocencia, la libertad ambulatoria y la privacidad, entre otros derechos fundamentales que deben ser resguardados sin excepción.

Por estos motivos, requerimos que se amplíe la descripción del numeral a los fines de poder conocer qué tipo de categorización y/o identificación biométrica se ha previsto incluir como de alto riesgo en el diseño reglamentario, y exigimos la incorporación de los usos más problemáticos de estos sistemas entre aquellos de riesgo inaceptable.

8.2. Los sistemas de IA que sirven de apoyo a autoridades jurisdiccionales

Debemos detenernos brevemente en la redacción del artículo. 14. 2. g) "Apoyo en la toma de decisiones, interpretación normativa o de los hechos por una autoridad jurisdiccional."

Cualquier uso de sistemas de IA para asistir en el proceso de administración de justicia debe ser celosamente supervisado por personas naturales. Si bien es dable considerar estos sistemas como de alto riesgo, sugerimos ampliar la definición para incorporar salvaguardas específicas y adecuadas para aquellos usos de tecnologías de IA que automatizan, asisten o reciben la delegación de tareas en el marco de la prestación de un servicio tan crítico y sensible como es el de la justicia.

8.3. Los sistemas de IA usados para la evaluación de riesgo de delitos o infracciones penales o reincidencia de personas

En el artículo 14.2. se clasifica como de riesgo alto a los sistemas de IA que se utilizan en la "Evaluación de riesgo de delitos o infracciones penales o reincidencia de personas naturales y

de riesgo de victimización [...] elaborar el perfil de personas naturales o grupos sobre conductas delictivas pasadas o durante la detención, investigación o enjuiciamiento de delitos [...]".

El citado artículo es una clara vulneración a la presunción de inocencia al permitir el uso y desarrollo de sistemas de IA que creen perfiles delictivos a partir del análisis de "grandes volúmenes de datos para detectar patrones aplicables a personas humanas". ¿A partir de qué datos se elaborarían estos perfiles criminales? ¿características personales? ¿datos sensibles? ¿sesgos? ¿motivos prohibidos de discriminación? Una permisión como esta supondría adelantar la sentencia y criminalizar ciertos perfiles de personas, antes del proceso penal.

Sumado a ello, un sistema de IA dedicado a tal finalidad vulnera además el principio de reeducación, rehabilitación y reincorporación social como fin de la pena, conforme prescribe la Constitución (art. 139, inciso 22). Debido a que mantiene un sistema de vigilancia y monitoreo sobre las personas privadas de libertad que han cumplido su sanción penal.

Resulta sumamente grave además que se permita para esta finalidad el tratamiento de bases de datos personales y sensibles, que pueden provenir de cualquier fuente masiva de datos, sea pública o privada, contenga información verídica o adulterada. Cabe recalcar que a la fecha para la intervención en sistemas informáticos o comunicaciones privadas se requiere una orden judicial. No se pueden esquivar estas garantías para proponer el uso y desarrollo de un sistema cuya finalidad implique la renuncia al debido proceso y a derechos fundamentales. Por ello, los supuestos contenidos en este art. 14.2.i deberían ser calificados como inaceptables.

9. Críticas a la lista de sistemas considerados de riesgo inaceptable

Es momento de centrarnos en uno de los aspectos más trascendentales de la propuesta reglamentaria, esto es, sobre los riesgos inaceptables y su consecuente prohibición. Comentaremos sobre los aspectos más problemáticos de la actual redacción, a los fines de apoyar su modificación para lograr una configuración que sea más comprensiva de las garantías necesarias para resguardar derechos fundamentales, más allá de las consideraciones que hemos presentado en el apartado 7 del presente documento.

9.1. Los sistemas de IA que pretenden modificar comportamientos humanos

En primer lugar, el Art 15.1. Señala como inaceptable aquellos sistemas que tienen como propósito "Modificar el comportamiento de una persona a través de técnicas subliminales o técnicas deliberadamente engañosas o el aprovechamiento de vulnerabilidades de un grupo poblacional específico cuando sea probable que su uso genere perjuicios considerables a una persona natural o a la colectividad."

No hay forma de trascender la conciencia de una persona sin que esto sea una grave violación de sus derechos fundamentales, incluido el derecho a la libertad de pensamiento, de conciencia, de religión y de dignidad humana. El proyecto debería prohibir completamente el uso de todo sistema de inteligencia artificial que "se sirva de técnicas subliminales que

trascienden la conciencia", independientemente de si estos puedan causar daño físico o psicológico. Por otro lado, la condición adicional de que tal distorsión deba realizarse de una manera que cause o sea probable que cause "perjuicios físicos o psicológicos" sugiere, erróneamente, que el comportamiento de una persona podría distorsionarse de una manera beneficiosa, lo cual es irrelevante frente a la contradicción que supone para los ya mencionados derechos fundamentales.

9.2. Los sistemas de IA que se emplean para la cuantificación de la conducta social de una persona o colectivo

Por su parte, los artículos 15. b) y c) señalan los sistemas con el fin de "Clasificar la fiabilidad de una persona natural o colectivo a través de la cuantificación de su conducta social, siempre que dicha clasificación le genere un efecto perjudicial y el contexto social no guarde relación con los datos de entrada" y los sistemas para "Clasificar la fiabilidad de una persona natural o colectivo a través de la cuantificación de su conducta social si le genera un efecto perjudicial y resulta injustificado o desproporcionado con respecto a su conducta social."

Mismo argumento puede presentarse en contra de la actual redacción del artículo 15. b) y c), sobre scoring social puesto que la mera intención de clasificar la fiabilidad de las personas es contraria a los derechos fundamentales, más allá de que esto pudiese resultar en efectos perjudiciales o resulte desproporcionado.

Sugerimos modificar el artículo 15.1 para que quede redactado de la siguiente manera.

> Artículo 15. Sistemas basados en inteligencia artificial de riesgo inaceptable
>
> 15.1 Un sistema basado en inteligencia artificial es considerado de riesgo inaceptable cuando se utiliza para:
>
> a) Sistema de IA que se sirva de técnicas subliminales que trascienden la conciencia de una persona incluyendo cuando dichas técnicas trascienden el comportamiento de la persona.
> b) Sistema de IA que aproveche alguna de las vulnerabilidades, ya sea intencionalmente o no, de un grupo específico de personas debido a su edad o discapacidad física o mental, o por cualquier otra razón que pudiera ocasionar discriminacion conforme.
> c) Sistemas de IA con el fin de calcular o establecer un récord de crédito social a partir de la evaluación o clasificación de personas físicas atendiendo a sus características físicas, sociales o a sus características personales o de su personalidad conocidas o predichas.
> d) Cualquier implementación y uso de sistemas de IA para el reconocimiento automatizado de rasgos humanos en espacios públicos, incluyendo el reconocimiento facial, la forma de caminar, las huellas dactilares, el ADN, la voz, y cualquier otras

señales biométricas, fisiológicas o de comportamiento independiente del fin que se persiga.

Nótese que hemos agregado un numeral más a este listado, para prohibir el reconocimiento automatizado de personas en espacios públicos a partir de sus datos biométricos en tiempo real, que se encuentran en clara contradicción con los derechos fundamentales. Estos sistemas son violatorios del principio de inocencia y generan un efecto inhibitorio en la sociedad que es desproporcionado para el fin que se persigue con su implementación.

10. Necesidad de incorporar estudios de impacto a derechos fundamentales

En el Proyecto normativo, el art. 24.1 afirma lo siguiente: “Se fomenta que los sistemas basados en inteligencia artificial se desarrollen bajo un enfoque de gestión de riesgos durante todo su ciclo de vida pudiendo incorporar medidas como pruebas piloto, sistemas de respaldo, auditorías algorítmicas, entre otros.”

Nos interesa detenernos brevemente sobre este artículo, simplemente para subrayar la necesidad de incorporar a los estudios de impacto a derechos fundamentales como el mecanismo adecuado para llevar adelante la gestión de riesgos previo al despliegue de los sistemas. En este sentido, lo dicho en el punto 4 vale para este apartado. Otro mecanismo adecuado para la gestión de riesgos es la implementación de medidas de apoyo a la innovación, como pueden ser los “espacios controlados de experimentación” (o sandboxes en inglés). Los cuales suponen contextos donde es posible llevar adelante procesos de desarrollo, prueba y validación de sistemas innovadores de IA que utilicen o se sirvan de datos personales, mientras que se analizan los riesgos inherentes a su puesta a disposición y sus medidas de mitigación. El objetivo es apoyar los procesos de innovación a la vez que se sostienen los deberes de supervisión y protección de los datos personales. Para que estos contextos controlados funcionen es necesario la supervisión e intervención constante de las autoridades nacionales de protección de datos personales. Esta será la encargada de poner a disposición del proceso experimental los conjuntos de datos personales que habiendo sido legalmente recopilados con otros fines, serán utilizados por el responsable del sistema de IA sujeto a desarrollo o validación.

11. Sobre los tipos de riesgos medios y bajos

En el capítulo III, sobre la gestión de riesgos en los sistemas basados en IA, se mencionan cuatro niveles de clasificación de riesgos: inaceptable, alto, medio y bajo. No obstante, si bien se enumeran y desarrollan los tipos de riesgos altos (art.14.2) e inaceptables (art.15) no existe mayor referencia en el capítulo o en el Proyecto en general sobre los tipos de riesgos medios y bajos. En otras palabras, no se enumeran los supuestos que calificarían como riesgos medios o bajos, ni las consecuencias u obligaciones que se derivarían de calificar a un riesgo como tal.

Por ello, en caso de sostener el diseño basado en matriz de riesgos, sería necesario incluir estas disposiciones.
En esta misma línea, es necesario incorporar referencias claras que permitan la clasificación de sistemas que actualmente no se encuentran previstos en la clasificación propuesta.Debido a que, por un lado, esto otorga resiliencia normativa el proyecto reglamentario, y por el otro, abstrae la necesidad de identificar riesgos, que ya no se sujeta a las funciones que cada sistema tiene, sino más bien a los riesgos que estos pueden presentar independientemente de los supuestos beneficios que se alcanzan con su despliegue.

III. Sugerencias de redacción

Finalmente, en atención al orden y coherencia que debe caracterizar a cualquier sistema jurídico, sobre todo en un tema de difícil acceso o con un alto nivel de complejidad para la ciudadanía, es que presentamos cinco sugerencias de redacción.

En primer lugar, tomando en cuenta el artículo 3, de definiciones, se sugiere redefinir algunos conceptos con las siguientes modificaciones:

- Poblacion vulnerable: Aquella conformada por personas o grupo de personas que estan expuestas a cualquier riesgo, desprotección familiar o discriminación. Dichas personas o grupo de personas pueden ser impactados por las brechas digitales existentes.

Por otro lado, en cuanto a la definición del concepto de IA o de sistema de IA propuesta en el Proyecto, se tiene que, definir a la inteligencia artificial y a los sistemas de IA aumenta la confusión, lo cual es justamente lo opuesto a lo que un artículo de definiciones debería hacer. La inteligencia artificial no se encuentra separada de aquel soporte que la contiene y que permite eventualmente llevar adelante su función (es decir, en tanto sistema). Además, la IA supone una disciplina de la técnica y como tal está sujeta al desarrollo de métodos, procesos y soluciones técnicas. En este sentido, no existe por fuera del estado actual o futuro de la técnica y por ende no es comparable con la inteligencia orgánica o “natural”, que eventualmente puede prescindir de una solución técnica para manifestarse.

En consecuencia, debemos señalar que el concepto de Inteligencia Artificial surge de un intento claro y manifiesto de obtener apoyo para el desarrollo de una específica disciplina de la técnica (con origen en Dartmouth College, en New Hampshire, Estados Unidos, 1955) y que por ende el vocablo “inteligencia” no hace referencia a la capacidad de emular las características específicas de lo “inteligente”, sino más bien a la búsqueda constante de la ciencia por comprender los misterios de esta y eventualmente aprovechar estos hallazgos para traducirlos en sistemas de base computacional.

Por este motivo, recomendamos definir el concepto de IA o de sistema de IA (cualquiera de los dos, pero sólo uno), de la siguiente manera:

- Inteligencia Artificial: “Sistema computacional que puede, para un conjunto determinado de objetivos, generar resultados tales como predicciones, recomendaciones o decisiones que influyen en entornos reales o virtuales. Los sistemas de IA pueden estar diseñados para operar con distintos niveles de autonomía”[12].

En segundo lugar, resulta recomendable incluir dentro del principio de múltiples partes interesadas a la comunidad técnica (art. 7.3, art. 9.2, art. 20.3, art. 20.5 y art. 32.3), tal como lo hace, a nivel nacional, la Resolución que crea la Alianza Nacional por una Internet Segura[13], en cuyo artículo 1 se menciona a la comunidad técnica entre los actores encargados de promover la participación activa.

Asimismo, a nivel internacional, se reconoce también la importancia de la comunidad técnica en el ámbito de derechos digitales en el *zero draft* del Global Digital Compact[14]. En este Pacto Digital de Naciones Unidas se reconoce el principio de multi-stakeholder o múltiples partes interesadas, entre las cuales se encuentran, -además del sector privado, sociedad civil y academia-, la comunidad técnica. De esta forma, es imprescindible, dentro del panorama de transformación digital y especialmente cuando se habla de inteligencia artificial, asegurar la presencia de la comunidad técnica.

En tercer lugar, en el artículo 8.2.a. se señala: “Para fortalecer la gobernanza en el uso y desarrollo de la inteligencia artificial, las entidades públicas son responsables de las siguientes obligaciones [...] a. Coordinar con otras entidades públicas que administran información oficial del Estado [...]”. En este artículo sería preferible colocar únicamente “información” y no “información **oficial**”.De lo contrario, la redacción podría inducir a creer que existen entidades públicas que administran información no oficial.

Por último, por una cuestión de política legislativa y escritura jurídica, sería recomendable que el artículo 14 (clasificación de los sistemas basados en IA) sea un artículo independiente, separado del art. 14.1. En el Proyecto se encuentran unidos, pese a que existen ideas diferentes en el mismo párrafo: los criterios de clasificación, el órgano competente para elaborar los lineamientos de evaluación y clasificación de riesgo, y las obligaciones del implementador. Sumado a ello, existe un subtítulo para la lista de riesgos inaceptables (art 15), pero no para la lista de riesgos altos (art. 14.2). Por lo que se recomienda que se señale un subtítulo propio para la lista de riesgos altos.

IV. Agradecimiento

[12] Definición adaptada a partir de la ofrecida por el National Institute of Standards and Technology, del Departamento de Comercio de Estados Unidos. (Enero de 2023) Artificial Intelligence Risk Management Framework.
Disponible en https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
[13] Secretaría de Gobierno y Transformación Digital (2024). Resolución 001-2024-PCM/SGTD.
[14] Naciones Unidas (2024). Global Digital Compact: zero draft.

Por último, agradecemos la oportunidad de presentarle nuestras recomendaciones y le solicitamos por este medio tenga a bien incluir nuestra participación en las sesiones en las que se discuta el Proyecto. Quedamos a vuestra disposición para colaborar en el desarrollo de una reglamentación que permita una regulación coherente con los derechos humanos de la ciudadanía peruana.

Cordialmente,

Franco Giandana Gigena

Access Now

Rubiela Gaspar Clavo

Hiperderecho