IA Ciudadana considera que el Reglamento de Inteligencia Artificial no ha logrado un estándar adecuado de protección de los derechos humanos

Desde IA Ciudadana, coalición de 17 organizaciones sociales liderando en España la respuesta social a las políticas de regulación e implementación de la inteligencia artificial, hemos llevado a cabo un primer análisis del texto final del Reglamento de Inteligencia Artificial (Reglamento de IA) aprobado en la Unión Europea[1].

Recordamos que, como coalición, una de nuestras principales peticiones es que la **sociedad civil participe en la definición de las políticas de inteligencia artificial.** La transparencia y la participación son requisitos básicos en toda democracia. Para ello, es indispensable que la ciudadanía disponga de información accesible y transparente acerca de las funciones, agendas y mecanismos de participación en los organismos responsables de regular y controlar la aplicación de la inteligencia artificial. El conjunto de la ciudadanía tiene derecho a participar de los asuntos públicos que, como los sistemas de IA, tienen y tendrán impacto en sus vidas, sus derechos y sus oportunidades. La participación pública efectiva y en plena diversidad permite decisiones más equitativas y con mayor legitimidad democrática. En el caso de España, según la reciente Estrategia de IA 2024 del Gobierno, sería la Agencia Española de Supervisión de la IA (AESIA) la encargada de articular un diálogo entre la comunidad científica, la industria y la sociedad civil, así como validar modelos de IA y el establecimiento de sus condiciones de evaluación, generando buenas prácticas, principios y recomendaciones. Sin

[1] El texto ha sido aprobado el pasado 21 de mayo por los miembros del Consejo de la Unión Europea. Resta como último paso su publicación en el Diario Oficial de la UE, lo que puede suceder en cualquier momento y tras lo cual entrará en vigor.

embargo, hasta el momento esta participación de la sociedad civil en la definición de políticas públicas de IA en España ha sido prácticamente inexistente.

Aunque habrá que ver cómo se aplica e interpreta este Reglamento, podemos adelantar que **no se han establecido mecanismos suficientes de participación ni se ha impulsado la transparencia necesaria para garantizar un control cívico de la utilización de estas tecnologías**. Para empezar, y como se explicará en detalle, aunque el reglamento exige la creación de una base de datos pública donde deberán registrarse los sistemas de IA, lo cierto es que esa obligación está limitada a algunos casos (proveedores y solo implementadoras que sean autoridades públicas). Además, esta base de datos no tendrá que ser pública para usos relacionados con las fuerzas de seguridad o en contextos migratorios (cuyos usos son de los más preocupantes respecto de los derechos humanos que podrían afectar). Por otro lado, el articulado no menciona la participación de la sociedad civil en obligaciones claves, como es la de llevar a cabo un estudio de impacto de derechos fundamentales. El Reglamento limita el escrutinio y control social en aspectos claves. Por ello, desde IA Ciudadana trabajaremos en **exigir mayor participación pública en políticas públicas de IA, en el diseño, la implementación y la supervisión de la IA**, y también en la **creación un registro público de algoritmos para lograr que España fomente la transparencia y la rendición de cuentas.**

Por su parte, en materia de prácticas que se consideran "inaceptables" dentro de la UE, es necesario resaltar que todas las prohibiciones que establece la norma contienen importantes y decepcionantes lagunas, lo que significa que no alcanzarán su máximo potencial. En algunos casos, estas lagunas podrían tener el efecto contrario al de prohibición: lanzan el mensaje de que algunas formas de vigilancia biométrica masiva y de discriminación impulsadas por IA son legítimas en la UE, lo que corre el riesgo de sentar un peligroso precedente global. A su vez, destacamos que, como resultado del intenso lobby de la industria, el Reglamento de IA presenta importantes deficiencias en aspectos clave, como lo es la clasificación de riesgos, que dificultarán su interpretación y, por ende, su aplicación efectiva.

Por último, debe destacarse que, al tratarse de una norma armonizada, deja muy poco margen a los Estados miembros para dictar normas más garantistas. Sin embargo, **desde IA Ciudadana hemos identificado la posibilidad de dictar normas más protectoras en materia de reconocimiento biométrico en espacios de acceso público**, por lo que seguiremos luchando porque en España se establezca una prohibición total de este tipo de tecnologías (incluido el **reconocimiento de emociones**), que consideremos van en contra de los principios democráticos y de los derechos reconocidos en un Estado Social y Democrático de Derecho.

Aclarado ello, en este documento hemos recogido las demandas que desde IA Ciudadana[2] proclamamos para proteger los derechos fundamentales en el Reglamento de IA y las hemos

[2] "60 organizaciones de la sociedad civil instan a la Presidencia española a fortalecer la protección de los derechos fundamentales en el Reglamento de Inteligencia Artificial de la Unión Europea" (mayo de 2023), https://cecu.es/notas/60-organizaciones-de-la-sociedad-civil-instan-a-la-presidencia-espanola-a-fortalecer-la-proteccion-de-los-derechos-fundamentales-en-el-reglamento-de-inteligencia-artificial-de-la-union-europea/

comparado con el texto final (valorando según lo pedido y lo recogido en la norma), lo que nos permite afirmar que no se ha logrado un estándar adecuado de protección de los derechos fundamentales[3]. Veamos:

Criterios de valoración:
Rojo (insuficiente), celeste (neutral) y **verde (bien)**

1. Ampliación del ámbito de aplicación del Reglamento

Se ha adoptado una definición limitada de sistema de IA (art. 3.1):

Desde IA Ciudadana requerimos una definición amplia de sistemas de IA que incluyera enfoques de aprendizaje automático, basados en la lógica, el conocimiento o estadísticas.

En la versión acordada, la definición ha sido limitada a *"(...) un sistema basado en una máquina diseñado para funcionar con diferentes niveles de autonomía y que pueden mostrar adaptabilidad después del despliegue y que, por razones explícitas u objetivos implícitos, infiere, a partir de los insumos que recibe, cómo generar resultados, tales como predicciones, contenidos, recomendaciones o decisiones que puedan influir física o entornos virtuales"*.

Al incluirse en la definición a sistemas con "*niveles de autonomía*" y que "*pueden exhibir adaptabilidad después del despliegue"* se limita el alcance del Reglamento porque es una característica intrínseca a algunos, pero no a todos los sistemas de IA. Esto dejaría afuera de la regulación a los sistemas de IA basados en reglas (que siguen un algoritmo predeterminado sin la capacidad de auto aprender y adaptarse a partir de nuevas entradas de datos después

Carta no pública enviada a la entonces presidencia española del Consejo de la Unión Europea titulada "Recomendaciones IA Ciudadana de cara al próximo trílogo sobre el Reglamento de IA" (julio de 2023).

"Declaración IA Ciudadana sobre la regulación europea de inteligencia artificial" (octubre de 2023), https://iaciudadana.org/2023/10/05/declaracion-de-ia-ciudadana-sobre-la-regulacion-europea-de-la-inteligencia-artificial/

[3] Análisis: La Ley de IA de la UE fracasa en garantizar la protección de los derechos humanos, https://algorights.org/la-ley-de-ia-de-la-ue-fracasa-en-garantizar-la-proteccion-de-los-derechos-humanos/

del despliegue) y que son los que más se utilizan hoy en día[4], con demostrada evidencia de su impacto en la vulneración de derechos[5].

2. Ampliación de la lista de prácticas prohibidas

Desde IA Ciudadana pedíamos que no se permitieran tecnologías que van en contra de los derechos fundamentales, como el reconocimiento biométrico en espacios de acceso público, el reconocimiento de emociones, el perfilado, la policía predictiva, la evaluación y perfilado en fronteras, las tecnologías que puedan manipular a las personas. Sin embargo:

No se ha prohibido totalmente el reconocimiento biométrico en espacios de acceso público (ni en tiempo real ni ex post).

Desde IA Ciudadana se pedía la prohibición total del reconocimiento biométrico remoto por parte de autoridades públicas y entidades privadas en espacios de acceso público, tanto en tiempo real como ex post. Ello, por considerar podría dar lugar a la vigilancia masiva biométrica generalizada, afectando el derecho a la privacidad, entre otros.

Lamentablemente, en el texto acordado la prohibición se limitó al reconocimiento biométrico en espacios de acceso público en tiempo real por parte de autoridades públicas de aplicación de la ley y, además, con tres excepciones relacionados con delitos graves.

Por su parte, el reconocimiento biométrico ex post no fue ni siquiera prohibido, sino que se encuentra permitido (como una práctica de alto riesgo) para autoridades de aplicación de la ley, con autorización judicial o administrativa previa.

No se ha prohibido totalmente el reconocimiento de emociones.

[4] Algunos ejemplos problemáticos conocidos que, en principio, quedarían fuera de esta regulación:

- El escándalo por los beneficios de cuidado infantil holandés (SYRI) consistía en una hoja de cálculo y un guión de programación relativamente simple para crear perfiles de riesgo. Sin embargo, arruinó miles de vidas.
- En España un ejemplo podría ser el software Bosco que determina quién tiene acceso a bono social de electricidad y respecto del cual la apertura de su código fuente se encuentra judicializada, siendo que tanto el Juzgado como la Audiencia Nacional lo han rechazado por supuestas razones de seguridad: https://civio.es/novedades/2024/05/08/la-audiencia-nacional-vuelve-a-rechazar-abrir-el-codigo-fuente-que-decide-quien-recibe-el-bono-social/
- El "algoritmo AMS" del Servicio Público de Empleo de Austria calculó futuras posibilidades de los solicitantes de empleo en el mercado laboral para decidir quién recibe más educación y formación. Para ello, utilizó regresiones logísticas simples. Sin embargo, discriminó a las mujeres al asignarles puntajes más bajos, incluso a pesar de que tenían la misma experiencia y calificaciones que los hombres.

[5] Access Now, Artificial Intelligence: what are the issues for digital rights?, https://www.accessnow.org/artificial-intelligence-issues-digital-rights/

Desde IA Ciudadana requeríamos que se establezca una amplia prohibición sobre el reconocimiento de emociones, porque suelen ser sistemas muy invasivos[6] y que presentan muchas inexactitudes y no gozan de evidencia científica[7].

En la versión final solo se prohibieron los sistemas de IA utilizados para inferir las emociones de una persona física en el lugar de trabajo y en las instituciones educativas, excepto cuando sea por razones médicas o de seguridad. Todos los demás usos del reconocimiento de emociones son de alto riesgo.

No se ha prohibido totalmente el uso de herramientas de policía predictiva.

Desde IA Ciudadana pedíamos que las autoridades administrativas, policiales y judiciales no pudieran usar sistemas de IA para hacer predicciones, perfiles o evaluaciones de riesgo para predecir delitos, porque afecta derechos fundamentales, como la tutela judicial efectiva o el principio de inocencia, y tienden a afectar a población racializada y en situación de vulnerabilidad[8].

Finalmente, se prohíbe el uso de sistemas de IA para realizar evaluaciones de riesgo de personas físicas con el fin de estimar o predecir el riesgo de que cometa un delito, basándose únicamente en su perfilado o en la evaluación de sus rasgos y características de su personalidad. Y con la excepción de sistemas que apoyen la evaluación humana de la implicación de una persona en un delito.

No se prohíbe ni la evaluación ni el perfilado en contexto migratorio.

Desde IA Ciudadana se pedía prohibir sistemas de evaluación y perfilado de riesgos individuales basados en la IA en el contexto de la migración; el análisis predictivo para interdicción, restricción y prevención de la migración; y los polígrafos de IA para prevenirla. Ello, por considerar que puede afectar seriamente derechos fundamentales, como el derecho a no ser discriminado[9].

Lamentablemente, bajo el Reglamento se permiten los sistemas de IA en contextos migratorios, incluida la biometría.

Se prohíbe la categorización biométrica para rastrear y juzgar personas en espacios públicos, pero de forma confusa y poco clara.

Se introdujo una nueva prohibición de los sistemas de categorización biométrica que infieren atributos sensibles (raza, opinión política, orientación o vida sexuales, sindicato, creencias religiosas o filosóficas), excepto el etiquetado o filtrado de conjuntos de datos biométricos

[6] Sánchez-Monedero, J., & Dencik, L. (2022). The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl. *Information, Communication & Society*, *25*(3), 413–430. https://doi.org/10.1080/1369118X.2020.1792530

[7] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements. Psychological Science in the Public Interest, 20(1), 1-68. https://doi.org/10.1177/1529100619832930

[8] Fair Trails, Report: Automating Injustice, https://www.fairtrials.org/articles/publications/automating-injustice/

[9] Algorace, Una introducción a la IA y a la discriminación algorítmica para movimientos sociales, https://www.algorace.org/wp-content/uploads/2023/10/informe-algorace-2.pdf

adquiridos legalmente, como imágenes, basados en datos biométricos o la categorización de datos biométricos en el área de aplicación de la ley.

Técnicas subliminales que causan distorsión del comportamiento y daño.

Originalmente esta prohibición hacía referencia a sistemas de IA que tuvieran el "propósito" de causar un "daño físico o psicológico". En el texto acordado esta prohibición se amplió al referirse al despliegue de técnicas subliminales más allá de la conciencia de una persona o técnicas deliberadamente manipuladoras o engañosas, que tengan el objetivo o "el efecto" de distorsionar materialmente el comportamiento de una persona, causando un daño significativo.

Sistemas de IA que afecten grupos vulnerables.

En la propuesta original esta prohibición estaba enfocada a determinados grupos vulnerables, como la infancia o las personas con discapacidad. Y también hacía referencia a sistemas de IA que tuvieran el "propósito" de causar un "daño físico o psicológico"

En la versión final se hace referencia a sistemas de IA que exploten vulnerabilidades relacionadas con la edad, la discapacidad o la situación social o económica específica con el objetivo o el efecto de distorsionar el comportamiento de alguien de una manera que cause un daño significativo.

3. Eliminación de cualquier tipo de discrecionalidad en el proceso de clasificación de sistemas de alto riesgo y ampliación de la lista del Anexo III

Clasificación de sistemas de alto riesgo.

Inicialmente, todos los casos de uso incluidos en la lista de alto riesgo tenían que seguir obligaciones específicas. Sin embargo, como resultado del intenso lobby de la industria, ahora los proveedores podrán decidir si sus sistemas son de alto riesgo o no, como un "filtro" adicional para ese sistema de clasificación.

Desde IA Ciudadana hemos pedido que se establezca un claro sistema de clasificación de riesgos, sin lagunas ni discrecionalidad por parte de las empresas. De esto dependía la efectividad de la norma, ya que sobre los sistemas de alto riesgo recaen todas o la mayoría de las obligaciones que impone.

Sin embargo, el art. 6 refiere a que los sistemas enumerados en el Anexo III (listado de alto riesgo) no se considerarán de alto riesgo si no suponen un riesgo significativo de daño, a la salud, a la seguridad o a los derechos fundamentales de las personas naturales, incluso sin influir materialmente en el resultado de la toma de decisiones. Y explica que tal será el caso

si cumple uno o más de los siguientes criterios: (a) el sistema de IA está destinado a realizar una tarea procesal limitada; (b) el sistema de IA está destinado a mejorar el resultado de una actividad humana previamente realizada; (c) el sistema de IA está destinado a detectar patrones de toma de decisiones o desviaciones de patrones anteriores de toma de decisiones y no pretende reemplazar ni influir en la evaluación humana realizada previamente, sin una revisión humana adecuada; o (d) el sistema de IA está destinado a realizar una tarea preparatoria para una evaluación relevante para la finalidad de los casos de uso enumerados en el anexo III.

El proveedor que considere que su sistema no es de alto riesgo debe documentar esta autoevaluación de manera previa a que el sistema sea puesto en el mercado o en el servicio. Y deberá registrarse en la base de datos.

Por su parte, los sistemas que realicen perfilado o *profiling* se considerarán de alto riesgo siempre.

Más allá de ello, desde IA Ciudadana consideramos que el texto acordado presenta varias lagunas y fallas desde la perspectiva de protección de las personas al permitir tal evaluación y no establecer grandes criterios de supervisión.

Listado de sistemas de IA de alto riesgo del Anexo III.

El Reglamento de IA se focaliza en gran medida en los sistemas que se consideran de "alto riesgo" (sobre los que establece obligaciones especiales como la gestión de riesgos, gobernanza de datos, documentación técnica, transparencia e información para el implementador, supervisión humana, etc.), dejando a muchos sistemas de IA prácticamente sin regulación. En el texto acordado la lista de sistemas de alto riesgo se ha ampliado, aunque no lo suficiente.

4. Obligaciones significativas de rendición de cuentas y transparencia pública sobre los usos públicos de los sistemas de IA y sobre todos los " implementadores" de IA de alto riesgo

Desde IA Ciudadana se requería la obligación de todos proveedores, los implementadores (es decir, quienes usan el sistema) y de las autoridades, o de quienes actúen en su nombre, de registrar todos los usos de los sistemas de IA en una base de datos pública.

Obligación limitada de registrar sistemas de IA en base de datos pública.

Lamentablemente, la obligación de registrar los sistemas de IA quedó limitada a: (a) proveedores de sistema de alto riesgo (menos de infraestructuras críticas); (b) proveedores que se hayan autoevaluado como de no alto riesgo; (c) implementadores que sean únicamente autoridades públicas o en su nombre. Y en el caso de sistemas de IA en biometría,

aplicación de la ley y migración el registro será en una base de datos **NO pública** y la información será limitada. Esto supondrá que ni las personas afectadas, ni la sociedad civil, periodistas o académicos podrán ejercer el escrutinio público en estos ámbitos, que son proclives a violaciones de derechos fundamentales, ni podrán exigir responsabilidades.

Además, puede verse con claridad que los implementadores de sistemas de alto riesgo en el sector privado no estarán obligados a registrarlos en la base de datos — otra cuestión crítica.

5. Requerimiento de un estudio de impacto de derechos fundamentales para todos los sistemas de IA de alto riesgo de forma previa a que se pongan en marcha

Estudio de impacto de derechos fundamentales (EIDF) con alcance limitado.

El Parlamento Europeo había introducido este requisito para todos los sistemas de IA de alto riesgo. Esta decisión fue apoyada desde IA Ciudadana. Aunque este requisito se mantiene en el texto final, se ha visto limitado. El EIDF solo deberá ser realizado por implementadores de sistemas de alto riesgo que sea organismos públicos u operadores privados que prestan servicios públicos y operadores que implementen únicamente los sistemas de IA de evaluación crediticia y de seguros que enumera el Anexo III. Además, los implementadores solo tendrán que mitigar los riesgos después de materializarse, es decir, cuando el daño ya haya sido causado, lo que resulta preocupante.

Por su parte, debe destacarse que el articulado finalmente no hace referencia involucrar a la sociedad civil y demás partes afectadas en la evaluación del impacto sobre derechos fundamentales que deben realizar.

Por último, los implementadores están obligados a registrar un resumen del EIDF en la base de datos que exige el Reglamento, salvo cuando se trate sistemas de alto riesgo que se utilicen en el marco de las fuerzas de seguridad y control migratorio. El público ni siquiera tendrá acceso a la mera información que una autoridad está utilizando un sistema de inteligencia artificial de alto riesgo. La información relacionada con el uso de IA en estos ámbitos solo se incluirá en un base de datos no pública, lo que limita gravemente la supervisión y el escrutinio constructivo. Se trata de un hecho muy preocupante ya que los riesgos para los derechos humanos, el espacio público y la legalidad son probablemente los más graves en estas dos áreas. Además, mientras los implementadores están obligados a notificar a la autoridad de vigilancia del mercado pertinente el resultado de su EIDF, existe una exención por “razones excepcionales de seguridad pública”. Es la misma excusa que se utiliza a menudo como justificación para llevar a cabo acciones de gestión policial y fronteriza desproporcionadas.

6. Las personas deben tener una serie de derechos reconocidos, tales como el derecho a acceder a la justicia y a la reparación, incluido el daño colectivo

Reconocimiento de derechos en favor de las personas.

Desde los inicios de los debates alrededor de este Reglamento, IA Ciudadana ha requerido el reconocimiento de derechos en favor de las personas. En el texto final:

- Se ha introducido el derecho a presentar a reclamación ante la autoridad pertinente. **Sin embargo, aún no está claro con qué eficacia las autoridades podrán hacer exigir el cumplimiento y responsabilizar a los infractores.**
- Se ha introducido el derecho a ser informado y recibir explicaciones del rol que ha tenido un sistema de IA (de los de alto riesgo) en un proceso de toma de decisiones que pueda tener efectos legales sobre una persona o afectarla de una manera que podría tener impacto sobre su salud, seguridad o derechos fundamentales. **Sin embargo, plantea dudas sobre la practicidad y la accesibilidad para obtener explicaciones significativas por parte de los implementadores. Además, la eficacia de estos mecanismos en la práctica es incierta, dada la ausencia de disposiciones como el derecho de representación de las personas físicas, o la capacidad de que las organizaciones de interés público presenten reclamaciones ante los órganos de supervisión nacionales.**
- Se ha incluido el Reglamento en la Directiva de Acciones de Representación, que permite los reclamos colectivos. Esta Directiva tiene por objeto garantizar que las personas consumidoras puedan proteger sus intereses colectivos en la UE a través de acciones de representación, que son acciones judiciales ejercitadas por entidades de representación (denominadas entidades habilitadas), que pueden incluir medidas resarcitorias.
- **Lamentablemente, se ha eliminado el derecho a acudir a vía judicial contra las autoridades nacionales de supervisión.**

7. Los estándares armonizados deben utilizarse únicamente para definir requerimientos técnicos, no para definir o aplicar principios legales o relacionados con los derechos fundamentales y su desarrollo debe garantizar mayor participación de la sociedad civil

Desarrollo de estándares técnicos.

Hay que tener en cuenta que la mayoría de las obligaciones que dispone el Reglamento dependen del desarrollo de estándares técnicos. Y que los sistemas de IA de alto riesgo y los

modelos de propósito general que estén de conformidad con estándares armonizados se presumirán que cumplen con la norma.

Por eso, desde IA Ciudadana ponemos de resalto que los procesos europeos de estandarización están dominados por la industria, por lo que hay que involucrar a otros actores de la sociedad civil en desarrollar estándares de IA[10], ya que, como hemos visto, esta tecnología puede influir en nuestros derechos fundamentales. Más allá de que seguimos considerando que los estándares técnicos no deben involucrarse en materia de derechos fundamentales, lo cierto en la práctica será muy difícil trazar esa línea. De ahí la necesidad de ampliar la pluralidad de voces en la elaboración de estándares técnicos.

[10] Ada Lovelace Institute, Discussion paper: Inclusive AI governance, https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/