COMENTARIOS:

DECRETO REGLAMENTARIO LEY 2489 DE 2025: “ENTORNOS DIGITALES SANOS Y SEGUROS PARA NIÑOS, NIÑAS Y ADOLESCENTES”.

Fundación Karisma

20 AÑOS Dejusticia

EL VEINTE

IUSR

FLIP FUNDACIÓN PARA LA LIBERTAD DE PRENSA

¿QUIÉNES SOMOS?

Somos dos organizaciones de la sociedad civil dedicadas a la defensa de los derechos humanos, Karisma, El Veinte, Dejusticia y Fundación para la Libertad de Prensa (FLIP). Karisma es una organización de la sociedad civil que trabaja para que las tecnologías digitales protejan y promuevan los derechos humanos y la justicia social, a través de investigación, capacitación e incidencia en áreas como la democratización del conocimiento, la participación cívica, la autonomía, la dignidad y la inclusión social, apoyada por sus laboratorios K+LAB de privacidad y seguridad digital y el de apropiación tecnológica.

Por su parte, El Veinte es una asociación de profesionales del derecho, periodistas y personas defensoras de derechos humanos dedicada a la defensa de la libertad de expresión, que estudia y combate el acoso judicial como forma de censura mediante el litigio estratégico y la reflexión académica, fortaleciendo las condiciones legales para el debate público en Colombia.

La Fundación para la Libertad de Prensa (FLIP) es una organización no gubernamental que defiende la libertad de expresión y promueve un clima óptimo para que quienes ejercen el periodismo puedan satisfacer el derecho de quienes viven en Colombia a estar informados.

Dejusticia es un centro de estudios jurídicos y sociales dedicado al fortalecimiento del Estado de Derecho y a la promoción de los derechos humanos en Colombia y en el Sur Global. Como centro de investigación-acción, su objetivo es promover la justicia social y ambiental. Para ello, realiza estudios rigurosos y propuestas sólidas de políticas públicas, lleva a cabo campañas de incidencia en foros de alto impacto y litigios de interés público, diseña e imparte programas educativos y de formación, y fortalece el tejido de la sociedad civil. Una de sus líneas de investigación, la de transparencia y derechos digitales, se enfoca en los impactos de las tecnologías digitales en los Derechos Humanos.

El Centro de Internet y Sociedad de la Universidad del Rosario (ISUR) es un espacio interdisciplinario que investiga, forma e incide sobre los impactos sociales de la tecnología desde una perspectiva de derechos humanos e interés público, promoviendo el empoderamiento tecnológico y la democratización del conocimiento en Colombia y Latinoamérica.

Fundación Karisma · EL VEINTE · 20 AÑOS Dejusticia · ISUR · FLIP Fundación para la Libertad de Prensa

PRIMERA PREOCUPACIÓN: EL PROCESO DE FORMULACIÓN DEL DECRETO NO HA GARANTIZADO LA PARTICIPACIÓN EFECTIVA Y SIGNIFICATIVA

El proyecto de decreto fue sometido a comentarios el 31 de diciembre, sin contar con mesas previas de participación, pese a que desde el mes de julio de 2025 fue sancionada la ley que se busca reglamentar. Es decir, la construcción del contenido fue unilateral y no permitió espacios efectivos de contribuciones antes de su redacción final. Como parte del proceso de reglamentación fue citada una reunión en la que solo se escuchará a 3 organizaciones de sociedad civil, cada una por 5 minutos. Esta fue convocada el 19 de enero a final de la tarde, para ser realizada el 21 de enero. No son claros los siguientes pasos en el proceso de reglamentación después de la reunión y la valoración de los comentarios que envíen distintos actores.

Esta serie de hechos no permiten asegurar nuestro derecho político a la participación ciudadana en la toma de decisiones al interior de las corporaciones públicas (art. 40-5 Constitución Política). Su efectividad depende de la convocatoria desde la etapa de planeación, así como de la disponibilidad de información suficiente y oportuna sobre el contenido y el procedimiento a seguir en el proceso de formulación de políticas (Ley Estatutaria 1757 de 2015, art. 102).

De ahí que sea necesario, en lo sucesivo, que se nos informe la forma en la que continuará el proceso, los actores involucrados y la forma en la que se realizará el seguimiento una vez aprobada la política.

SEGUNDA PREOCUPACIÓN: IMPLEMENTACIÓN DE UN MODELO BASADO EN RIESGOS QUE DESPLAZA A UNO QUE SE BASE EN DERECHOS Y RIESGOS DE CENSURA INDIRECTA Y AFECTACIÓN A LA LIBERTAD DE EXPRESIÓN DERIVADOS DEL ETIQUETADO DE CONTENIDOS

Los sistemas basados en riesgos han sido utilizados como aproximación en la regulación en la Unión Europea para servicios digitales y para Inteligencia Artificial, también en las Directrices para la gobernanza de plataformas digitales. Su adopción por esas entidades no hace que sean apropiadas para el contexto colombiano, por muchas razones. La aproximación de riesgos, en temas de regulación de plataformas, supone el etiquetado y la clasificación de discursos. En la práctica, la clasificación de discursos supone desconocer que la libertad de expresión ampara los contenidos que no son ilegales para crear nuevas categorías de discursos prohibidos, en contravía de la Constitución, la Convención Americana y el Pacto de Derechos Civiles y Políticos. A su vez, para evitar

sanciones, se crea un incentivo para los sujetos de las obligaciones de sobre-moderar contenido que está amparado por la libertad de expresión.

Aunque el proyecto contempla la existencia de una "justificación legítima, proporcional y documentada" como una excepción frente a los parámetros más restrictivos, esta previsión resulta insuficiente para satisfacer los principios de legalidad y proporcionalidad exigidos por el test tripartito cuando se trata de limitar la libertad de expresión. Esta iniciativa recurre a conceptos indeterminados, ambiguos y abiertos, tanto para definir contenidos sujetos a restricción como para delimitar los supuestos que activan la intervención estatal y de las propias plataformas.

Las categorías utilizadas, tales como contenidos "ilegales", "cualquier tipo de violencia", "tratos inadecuados en línea"[1] o "contenido inapropiado", carecen de definiciones claras, objetivas y verificables impidiendo que los usuarios, creadores de contenido, periodistas, artistas e incluso que las mismas plataformas puedan prever con un grado razonable de certeza qué expresiones están prohibidas y cuáles se encuentran amparadas por la Constitución Política. En este contexto, expresiones protegidas, como manifestaciones artísticas, discursos críticos, sátira, contenidos informativos o expresiones de protesta, pueden ser subsumidas indebidamente dentro de categorías restrictivas, generando censura.

En los artículos 2.2.3--.7 y 2.2.3--.8, por ejemplo, se establece un sistema de clasificación y etiquetado de contenidos basado en categorías amplias de riesgo asociadas al contenido, la conducta, el contacto y el consumo. Estas categorías incluyen, entre otras, referencias a sexualidad, violencia, conductas peligrosas o contenidos potencialmente adictivos. La experiencia sobre moderación de contenidos demuestra que este tipo de clasificaciones amplias, especialmente cuando se implementan mediante herramientas automatizadas, tienden a generar fenómenos de sobre-moderación, despriorización preventiva de contenidos lícitos, puede operar como un mecanismo de deslegitimación o estigmatización del discurso, afectando su circulación y recepción aún cuando se trate de expresiones protegidas

Este riesgo es particularmente relevante para contenidos de carácter educativo, informativo o periodístico, como aquellos relacionados con educación sexual integral, salud mental, derechos humanos o participación social, que pueden ser etiquetados negativamente pese a su valor social. Aunque el parágrafo 1 del artículo 2.2.3--.7 reconoce que ciertos contenidos sensibles pueden tener un propósito educativo o informativo, el decreto no establece salvaguardas operativas suficientes que impidan que el etiquetado derive en bloqueos, restricciones de visibilidad o eliminación automática. Esto puede generar un efecto inhibidor sobre la libertad de expresión y fomentar la autocensura por parte de creadores y usuarios. A su vez, la carga se trasladaría a los usuarios que tendrían que enfrentar procesos particulares para demostrar que su contenido está cubierto por la excepción de contenido educativo, informativo o de prevención, y no que se presume que sus contenidos son expresiones libres.

[1] Pese a que se intenta definir, tampoco resulta clara en tanto que no describe la conducta, el daño ni los elementos que permiten identificarla sino que remite de manera genérica a otra categoría igualmente amplia referida a la taxonomía de riesgos 4C.

Resulta igualmente preocupante que, según el parágrafo del artículo .2.3--.7 será el MinTIC quien formule "lineamientos y protocolos para la respuesta frente a contenidos que conlleven a la generación de riesgos (...) con miras a obtener una respuesta oportuna por parte de las plataformas para su moderación o remoción y la coordinación con las autoridades respectivas". La formulación de la respuesta ante contenidos, a modo de sanción o de moderación o cualquier otra, debe ser de orden legal. Delegar al ministerio esa función, con una naturaleza legal de menor entidad contraría también la Constitución. Echamos de menos también, en ese sentido, que sea el mismo ministerio el plantee un modelo en el que no se tienen en cuenta a otros actores del ecosistema, pese a que el Estado colombiano asumió el compromiso de sostener un modelo de múltiples partes interesadas para la gobernanza de internet desde hace 20 años y que hace parte de la Mesa de Gobernanza de internet. Por lo mismo, tanto la formulación de políticas, como su implementación y la consiguiente rendición de cuentas deben estar guiadas por ese enfoque, que no fue incluido en la redacción del proyecto.

TERCERA PREOCUPACIÓN: RIESGOS PARA LA INTIMIDAD Y LA PROTECCIÓN DE DATOS PERSONALES EN EL PROCESO DE DETERMINACIÓN Y ASEGURAMIENTO DE EDAD

En los artículos 2.2.3--.2, 2.2.3--.5 y 2.2.3--.6, se desarrolla el diseño del denominado "Modo Niña, Niño o Adolescente" y la obligación de aplicar protecciones diferenciadas por grupos etarios presuponen la necesidad de determinar o estimar la edad de las personas usuarias. Sin embargo, el decreto se limita a exigir "mecanismos razonables de determinación de edad" sin definir su alcance ni establecer prohibiciones claras frente a métodos intrusivos. Esta ambigüedad normativa genera incentivos para que las plataformas adopten mecanismos de verificación de edad altamente invasivos, como el uso de documentos oficiales, datos biométricos o inferencias algorítmicas basadas en el comportamiento, con el fin de reducir riesgos regulatorios.

La experiencia comparada demuestra que estos sistemas conllevan riesgos significativos para la privacidad y la protección de datos personales. En Australia, en el marco de la Online Safety Act y los debates sobre age assurance, la eSafety Commissioner y diversos informes técnicos han reconocido que no existe actualmente un método de verificación de edad que sea simultáneamente preciso, no invasivo y respetuoso de los derechos fundamentales. La verificación de edad a gran escala ha sido criticada por favorecer la creación de bases de datos masivas de identidad, aumentar el riesgo de filtraciones de información y habilitar usos secundarios de los datos recolectados, particularmente problemáticos cuando se trata de niños, niñas y adolescentes.

De manera similar, autoridades europeas de protección de datos han advertido que los sistemas de verificación de edad pueden ser incompatibles con el principio de minimización de datos cuando existen alternativas menos intrusivas basadas en el diseño del servicio. Aunque el proyecto de decreto menciona la estimación anónima y la

minimización de datos, no excluye expresamente mecanismos intrusivos ni exige evaluaciones de impacto en protección de datos, lo que deja abiertas importantes brechas desde la perspectiva del derecho a la privacidad.

Sobre este asunto, la CNIL, autoridad francesa responsable de garantizar la protección de los datos personales, ha desarrollado seis principios esenciales para lograr compatibilizar la verificación de edad con la protección de los datos personales. Estos son: (i) proporcionalidad: que exige tener en cuenta los fines del tratamiento, destinatarios, tipos de datos recogidos, tecnologías disponibles y nivel de riesgo asociado al tratamiento; (ii) minimización: entendido como la limitación en la recolección de datos a lo estrictamente necesario y a la no conservación del dato una vez la verificación ha sido completada; (iii) robustez: exigiendo que los mecanismos de verificación de edad sean robustos para prácticas que representan grandes riesgos para NNA; (iv) simpleza: se debe fomentar el uso de soluciones sencilla y fáciles de usar; (v) estandarización: se sugiere adoptar un programa de certificación para garantizar el cumplimiento de estas normas; y (vi) intervención de terceros: con lo que se recomienda que sean terceros confiables quienes comprueben la edad.

Sobre el último principio la CNIL ha ahondado indicando que la intervención de terceros requeriría de dos partes. La primera, sería una en la que el tercero emite una prueba de edad y, la segunda, sería aquella en la que se transmite esa prueba certificada al sitio visitado para que este decida permitir o denegar el acceso al contenido. Este esquema busca enfrentar los riesgos propios de la privacidad al preservar el uso de Internet sin revelar la identidad del usuario, por ello busca la separación de funciones y de sujetos.

CUARTA PREOCUPACIÓN: IMPACTOS SOBRE EL ACCESO A INFORMACIÓN PLURAL EN EL DISEÑO SEGURO POR DEFECTO Y LIMITACIÓN DE SISTEMAS ALGORÍTMICOS

El decreto en el articulo 2.2.3--.3 impone obligaciones relacionadas con el diseño seguro por defecto, incluyendo la limitación de sistemas de recomendación y amplificación algorítmica para cuentas de niños, niñas y adolescentes. Si bien estas medidas buscan reducir la exposición a riesgos, su aplicación indiscriminada puede afectar la expresión y el acceso de adolescentes a información diversa y legítima, reforzando entornos excesivamente restrictivos que no fomentan el desarrollo de competencias críticas ni el ejercicio progresivo de la autonomía.

La simple reducción algorítmica de contenidos no sustituye políticas integrales de alfabetización digital y puede generar efectos adversos, como la creación de burbujas informativas hiper controladas o la invisibilización de contenidos relevantes para determinados grupos de adolescentes. Por ello, en caso de insistir en su implementación, sería determinante que se transparentaran los límites algorítmicos y de recomendación que resultaren impuestos, así como sus justificaciones; se garantizara la reducción

gradual de tales límites según la madurez de los NNA; y se implementaran políticas integrales de alfabetización digital.

QUINTA PREOCUPACIÓN: RIESGOS DE VIGILANCIA Y FALTA DE CONTROLES EN EL SISTEMA INTEGRAL DE MONITOREO Y EVALUACIÓN

El Sistema Integral de Monitoreo y Evaluación para la Protección de Niños, Niñas y Adolescentes en Entornos Digitales, establecido en el artículo 2.2.3--.11 plantea interrogantes relevantes desde la perspectiva de la privacidad, la libertad de expresión y la proporcionalidad. Aunque el artículo menciona la protección de datos personales, no establece límites claros sobre el alcance del monitoreo, los tipos de datos que pueden ser recolectados ni las finalidades específicas del tratamiento, pudiendo convertirse en una actividad de inteligencia. Tampoco establece responsabilidades específicas para cada uno de los actores.

La ausencia de salvaguardas robustas y de mecanismos de supervisión independiente puede derivar en prácticas de monitoreo amplio o análisis sistemático del comportamiento de niños, niñas y adolescentes en entornos digitales, lo que resulta incompatible con los principios de necesidad, finalidad y proporcionalidad en el tratamiento de datos personales. En todo caso y por las amplias facultades que tendría, el Sistema Integral de Monitoreo y Evaluación para la Protección de Niños, Niñas y Adolescentes en Entornos Digitales debería ser transparente y su funcionamiento debería ser de acceso ciudadano, pues no de otra manera se podría hacer veeduría y control para evitar abusos y extralimitaciones en el marco de nuestro sistema democrático.

SEXTA PREOCUPACIÓN: NECESIDAD DE REFORZAR EL PRINCIPIO DE PROPORCIONALIDAD Y LAS SALVAGUARDAS DE DERECHOS FUNDAMENTALES

En conjunto, las disposiciones analizadas evidencian el riesgo de que, en nombre de la protección de niños, niñas y adolescentes, se habiliten mecanismos de control, identificación y moderación que resulten desproporcionados y que afecten derechos fundamentales como la libertad de expresión, el acceso a la información y la privacidad.

La situación en Australia, Reino Unido y algunos estados de USA muestra que los enfoques centrados en verificación de edad y etiquetado amplio de contenidos deben

manejarse con extrema cautela, privilegiando soluciones basadas en el diseño del servicio, la minimización de datos y salvaguardas claras contra la censura y la vigilancia.

Por ello, resulta fundamental que el decreto incorpore criterios más estrictos y operativos de proporcionalidad, así como prohibiciones expresas frente a prácticas intrusivas, con el fin de garantizar que la protección reforzada de los derechos de niños, niñas y adolescentes no se traduzca en restricciones indebidas de otros derechos fundamentales.

SÉPTIMA PREOCUPACIÓN: INSUFICIENTE CONSIDERACIÓN DEL RIESGO QUE CONLLEVA EL DISEÑO ADICTIVO DE LAS PLATAFORMAS DIGITALES

El proyecto de decreto no tiene en cuenta suficientemente los riesgos asociados a los diseños adictivos de las plataformas y aplicaciones digitales que incentivan que NNA permanezcan cada vez más tiempo en línea. Aunque en el artículo 2.4 se menciona que se limitarán "funcionalmente aquellas características del servicio que puedan generar riesgos" de "exposición a dinámicas adictivas" y en las categorías temáticas se habla de "servicios adictivos", estas menciones no reflejan la magnitud ni la sistematicidad del problema.

Los estudios de psicología comportamental han documentado que redes sociales, videojuegos en línea, entre otras plataformas, incorporan deliberadamente características como sistemas de recompensas variables, desplazamiento infinito, notificaciones persistentes y personalización algorítmica que fomentan comportamientos en línea desregulados y adictivos. Esto puede ser especialmente riesgoso para los niños, pues estos están en una etapa en que aún no han desarrollado todas las partes de su cerebro completamente, así como tampoco han adquirido todas las habilidades emocionales.

En consecuencia, es conveniente que el decreto incorpore regulaciones adecuadas tendientes a mitigar el riesgo sistémico del diseño adictivo, quizás especificando que en el riesgo de consumo está el de la adicción.

OCTAVA PREOCUPACIÓN: LA APROXIMACIÓN DE RIESGOS SIN UNA POLÍTICA PÚBLICA ROBUSTA EN OTROS SECTORES CAUSARÍA SU INEFECTIVIDAD.

Alrededor del mundo se están implementando distintas formas de control del contenido de internet bajo el llamado a la protección de niñas y niños. Pese al poco tiempo de implementación, en distintos espacios se ha cuestionado su efectividad y el surgimiento de nuevos riesgos para la población que se pretende proteger: se ha incrementado la

descarga de servicios de VPN, que pueden poner en mayor riesgo los datos y la integridad de NNA. Se ha construido un mercado negro de identidades para sobrepasar los controles de identidad en Australia, llevando a NNA a caer en estafas para acceder a cuentas verificadas.

De cara a la adopción que han venido realizando grandes compañías de sistemas de identificación, el caso de Roblox, plataforma de juegos, ilustra con claridad los riesgos de implementar una política de verificación de edad centrada casi exclusivamente en una solución técnica aislada. La adopción de sistemas automatizados de estimación de edad, como el reconocimiento facial mediante inteligencia artificial, ha demostrado ser imprecisa y generar errores sistemáticos que clasifican incorrectamente a menores y adultos. Lejos de reforzar la protección, esta aproximación ha afectado negativamente la experiencia de uso y ha abierto la puerta a la aparición de mercados clandestinos e informales de venta de cuentas con verificación de edad comprobada para acceder a chats; así como al uso de VPS para evadir los controles.

Asimismo, la experiencia demuestra que una regulación que se limite a imponer mecanismos formales de verificación, sin analizar otros frentes ni las dinámicas reales de las plataformas digitales, corre el riesgo de ser ineficaz o incluso contraproducente. En este contexto, la regulación debería considerar que la verificación de edad no es un problema que pueda resolverse únicamente mediante una herramienta tecnológica, sino que requiere un enfoque más amplio, que tenga en cuenta limitaciones técnicas, impactos en derechos como la privacidad y efectos colaterales sobre el ecosistema digital.