

Los riesgos de la IA generativa para la protección de los menores

Como señala The Lancet en un reciente artículo¹, no deja de ser paradójico que las herramientas digitales diseñadas para educarnos, conectarnos y empoderarnos están suponiendo al mismo tiempo un factor de grave riesgo para la protección de la infancia y la juventud desde el punto de vista de su seguridad y su correcto desarrollo físico, mental y moral.

Cita la revista distintas fuentes, como el informe de la Sociedad Nacional para la Prevención de la Crueldad Infantil del Reino Unido (NSPCC) *Viewing Generative AI and children's safety in the round*², que pone de relieve dicho riesgo asociado a las herramientas de inteligencia artificial generativa. Señalando además cómo la IA puede facilitar la captación de menores, el acoso y la generación de contenidos perjudiciales para ellos. Muchos modelos AI generativa de código abierto están entrenados con material de abuso sexual infantil.

Se refiere también al hecho de que la explotación y el abuso sexual infantil en línea (OCSEA, por sus siglas en inglés) se ha convertido en una amenaza generalizada: según una revisión de Fry et al. publicada en The Lancet Child and Adolescent Health³, un elevado porcentaje de menores ha experimentado ese problema a lo largo de su vida, resaltando la urgente necesidad de intervenciones basadas en la evidencia. ¿Cómo podemos proteger a los niños y niñas en línea sin coartar su capacidad de aprender, socializar y crecer?, se pregunta The Lancet. Estos autores describen subtipos de OCSEA, como la solicitud de contacto en línea por parte de adultos para su captación, que afectaría aproximadamente a un 12,5 % de las personas menores de edad a nivel mundial.

El auge de las imágenes generadas por IA ha generado desafíos en la capacidad y las estrategias de respuesta. Un Informe sobre Abuso Sexual Infantil con IA de la Fundación Internet Watch reveló en 2023 que en tan solo un mes se subían a la dark web más de 11 000 imágenes relacionadas con el abuso y la pornografía infantil, y una actualización de julio de 2024 indicaba el agravamiento del problema⁴. El informe mostraba su

¹ [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(25\)00043-3/fulltext?dgcid=raven_jbs_etoc_email](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(25)00043-3/fulltext?dgcid=raven_jbs_etoc_email)

² <https://learning.nspcc.org.uk/research-resources/2025/generative-ai-childrens-safety>

³ *Prevalence estimates and nature of online child sexual exploitation and abuse: a systematic review and meta-analysis.* [https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642\(24\)00329-8/abstract](https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(24)00329-8/abstract)

⁴ <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/>

preocupación por el hecho de que el auge de esta modalidad y la necesidad de actuar contra ella suponga incluso detraer recursos que hoy se aplican a la lucha contra el material real de abuso y explotación sexual infantil.

Los Gobiernos están adoptando diferentes iniciativas ante esta preocupación y la alarma social que genera, en muchos casos no exentas de polémica con los expertos y la sociedad. En 2024, las autoridades australianas promulgaron la prohibición de las cuentas de redes sociales para menores de 16 años, algo que se ha planteado también en Europa (y España). La medida fue criticada por 140 académicos y profesionales de la salud mental⁵, que consideraban la medida "simplista" argumentando, entre otras cosas, que los contenidos online son fundamentales hoy para la formación de las personas menores de edad, y que la medida podría tener el efecto perverso de relajar las medidas de seguridad y protección por parte de las plataformas, o de desviar la navegación de niños, niñas y adolescentes hacia entornos menos controlados y seguros.

Estas posiciones matizadas recogen la preocupación por el tiempo excesivo que los menores pasan ante la pantalla, estimulados por las plataformas a través de los algoritmos y los llamados "patrones oscuros" que generan adicción, y también por su exposición a contenidos inadecuados y nocivos o por el ciberacoso. Pero destacan igualmente el papel positivo que desempeñan las redes en el fomento de las conexiones sociales, especialmente para los grupos marginados. Por ejemplo, un estudio realizado en 2022 reveló que las redes sociales pueden contribuir a la salud mental de los jóvenes LGBTQ+ mediante la conexión con sus iguales, la gestión de la identidad y el apoyo social⁶. Otro estudio reciente con más de 1200 alumnos de secundaria⁷ en Gran Bretaña, encontró poca diferencia en el bienestar mental de quienes asisten a escuelas con prohibiciones telefónicas restrictivas y escuelas sin ellas.

Entonces, si las prohibiciones generalizadas corren el riesgo de aislar a los niños de los beneficios del entorno digital ¿qué alternativas podrían ser más efectivas? Para The Lancet, se trata de priorizar la prevención más que las medidas reactivas.

⁵ https://westernsydney.edu.au/_data/assets/pdf_file/0016/2052160/Open_letter_re_social_media_bans.pdf

⁶ Social Media Use and Health and Well-being of Lesbian, Gay, Bisexual, Transgender, and Queer Youth: Systematic Review. <https://www.jmir.org/2022/9/e38449/>

⁷ *School phone policies and their association with mental wellbeing, phone use, and social media use (SMART Schools): a cross-sectional observational study*
[https://www.thelancet.com/journals/lanpe/article/PIIS2666-7762\(25\)00003-1/fulltext](https://www.thelancet.com/journals/lanpe/article/PIIS2666-7762(25)00003-1/fulltext)

Ello significa:

- Una obligación clara para las empresas tecnológicas de liderar la seguridad desde el diseño, exigiéndose a las redes sociales y a las plataformas medidas proactivas para detectar y eliminar contenido dañino.
- Desarrollo de marcos éticos para el uso de la IA que reconozca como prioridad la seguridad infantil.
- Un marco regulatorio sólido, diseñado con la participación de niños y padres.
- El fomento de la alfabetización digital, empoderando a los niños para que naveguen con seguridad en los espacios en línea.
- Endurecer las actuaciones contra los involucrados en la explotación de los menores en el entorno digital.

Centrarse en la prevención mediante la rendición de cuentas tecnológica, en regulaciones sólidas y en la alfabetización mediática e informacional (AMI), podemos proteger mejor a los niños de la explotación en línea y garantizar su seguridad en la era digital.