



El amor en tiempos de Grok

Inteligencia Artificial, sexismo real

¿Cómo le hace Cupido para flechar en 2026? Quizá le pregunta técnicas a GPT o estudia el algoritmo de compatibilidad entre sus objetivos. Seguramente el número de likes es decisivo para que evalúe las probabilidades de éxito de una pareja. ¿Apps de citas? Puede ser... pero ahí llega un problema difícil de sortear para su flecha milenaria: la foto quizá ya no es de quien ocupa el perfil, tal vez tampoco el video aunque pueda ver a la persona en movimiento y hasta escuchar su voz.

El amor en los tiempos de la Inteligencia Artificial (IA) abre una vulnerabilidad a la cual dedicamos este reporte mensual, a propósito del 14 de febrero: cuando la IA se usa para sexualizar y dañar la reputación, especialmente de niñas y mujeres, estamos ante una forma de violencia de género habilitada por esta tecnología. Y lamentablemente la legislación avanza menos rápido que los motores de entrenamiento de las grandes infraestructuras de IA.

No es accidente colateral: es una actualización técnica del patrón antiguo que convierte a las mujeres —sobre todo a las jóvenes— en objetos deseables y contenido replicable, incluso hasta hacer viral una imagen de la cual la víctima ni enterada está.



Caso Grok y sexualización sin consentimiento

A inicios de 2026, AI Forensics publicó un reporte sobre el uso de @Grok, la IA de X propiedad de Elon Musk, para generar imágenes sexualizadas. El equipo reunió 50 mil menciones a @Grok y 20 mil imágenes generadas entre el 25 de diciembre de 2025 y el 1 de enero de 2026. Ese periodo delimita el fenómeno y permite medir su escala con base en evidencia empírica. Además, esto documenta la facilidad de uso de los “prompts”, el potencial de viralidad y la grave ausencia de barreras eficaces. En términos feministas, es la misma violencia sexual de siempre, pero con una mejora operativa para los victimarios: se automatiza la degradación y se escala el daño.

En esas 20 mil imágenes, AI Forensics reportó que 53% incluyó personas con “vestimenta mínima”. Dentro de ese subconjunto, 81% fue clasificado como mujeres. Este patrón importa porque no describe la sexualización en general, sino una sexualización concentrada en las mujeres. No se trata de contenido neutral sino de la producción masiva de objetivación de las mujeres. **Cuando el resultado se concentra en mujeres, cabe la pregunta: ¿qué reproduce la IA del orden social todavía vigente? La respuesta apunta a una lógica patriarcal de género que normaliza una jerarquía en la que la mujer es el cuerpo disponible para el consumo visual.**

El análisis también registró una brecha de edad aparente entre las mujeres y los hombres retratados. Las mujeres aparecen con mediana de 22.5 años y los hombres con mediana de 31.5 años. Además, el 92 % de las mujeres retratadas aparentan tener menos de 30 años, frente al 40 % en los hombres. Estos datos conectan la sexualización con un ideal de juventud como criterio estético dominante. Si la sexualización se concentra en mujeres jóvenes, el sistema



machista refuerza un mandato cultural según el cual lo deseable es lo joven y lo controlable.

El riesgo crece cuando esta producción de imágenes se acerca a la niñez.

Con un clasificador de visión, AI Forensics estimó que 2% de las imágenes retrató a personas que aparentan 18 años o menos. En una muestra de 20 mil, ese porcentaje implica cientos de piezas con apariencia de menores, lo que eleva el daño y el umbral jurídico del problema. El umbral no es sólo moral: es legal y de protección de los derechos de las infancias. En términos de política pública, cuando una herramienta abre esa puerta —aunque sea en porcentajes pequeños—, habilita un mercado de depredación. **La estadística se vuelve tragedia porque cada punto porcentual corresponde a niñas, rostros, vidas y comunidades afectadas.**

El caso Grok no se reduce a que una IA genere imágenes. Importa que la generación ocurre dentro de X, un entorno con publicación inmediata, alcance algorítmico, búsqueda, interacción y archivo. Investigaciones periodísticas documentaron que usuarios utilizaron Grok para producir imágenes no consentidas de mujeres y contenido sexualmente violento, y que ese material circuló dentro de la misma plataforma, lo que acelera su reproducción y eleva el costo para las víctimas. La arquitectura de plataforma convierte el daño en un bucle: la víctima no puede salir del hecho porque este se vuelve buscable, descargable y reutilizable.

La escala también se explica por la dinámica viral del prompt. *The Guardian* describió cómo una tendencia inicial de “poner bikini” evolucionó en cientos de miles de solicitudes para retirar ropa de fotos de mujeres. Esa lógica convierte la agresión en una práctica repetible, socialmente incentivada y de bajo costo para el agresor. Esto tiene un nombre: cultura de impunidad. Si el agresor percibe que no pasa nada, la violencia se convierte en deporte de masas. Y si además existe una



recompensa social (“likes”, retuits...), el acto se reetiqueta como juego. El costo lo paga una sola parte: la mujer expuesta.

La reacción de la plataforma confirma el problema operativo. *The Guardian* reportó que, tras el escándalo, Grok desactivó la generación de imágenes para usuarios que no pagan por usar la plataforma y la restringió a cuentas de pago. El seguimiento periodístico añadió que existen variantes de Grok fuera del flujo público de @Grok, lo que complica el control cuando el modelo vive en más de un canal de acceso. **Cuando la mitigación llega después del escándalo, el mensaje es claro: primero se monetiza la atención, luego se gestiona el daño reputacional. El patriarcado y el capitalismo trabajando al unísono.**

México y el ciberacoso contra mujeres

Antes de fijar la atención en los casos virales, conviene recordar que en México existe violencia digital cotidiana y extendida. Por eso, cualquier herramienta que haga accesible al alcance de un clic la sexualización sin consentimiento cae en terreno fértil. En términos de prevención, esto cambia el enfoque, pues el riesgo no es hipotético, sino cotidiano. INEGI mide el ciberacoso mediante el Módulo sobre Ciberacoso. **En 2024, 21.0% de la población usuaria de internet de 12 años o más vivió ciberacoso, equivalente a 18.9 millones de personas.**

En mujeres usuarias de internet, la prevalencia fue 22.2%. Estos datos muestran que la violencia digital ya tiene una magnitud significativa entre la población y afecta a las mujeres con mayor frecuencia que al promedio general. Si millones ya viven ciberacoso, entonces herramientas como las de nudificación llegan a un ecosistema donde la agresión digital está normalizada. El desglose por





edad refuerza el foco en las mujeres jóvenes. En 2024, 31.1% de las mujeres de 20 a 29 años usuarias de internet reportaron ciberacoso en los últimos 12 meses. Este grupo coincide con el perfil de juventud que aparece en el caso Grok, que concentró la sexualización en mujeres con una apariencia más joven. Cuando la violencia se concentra en mujeres jóvenes, la edad es el lugar donde se cruzan deseo, poder, control y castigo. Y ese cruce es precisamente lo que la tecnología amplifica al permitir producir imágenes “por demanda”.

En MOCIBA 2024, varios estados de la República superan el promedio nacional en la prevalencia de ciberacoso en mujeres, mientras que otros quedan por debajo. Este contraste ayuda a ubicar focos de alerta territoriales y a exigir respuestas diferenciadas. Es clave no caer en el error de una respuesta única. La violencia digital se alimenta de factores locales: capacidades institucionales, cultura de denuncia, redes comunitarias, accesibilidad de los ministerios públicos, educación digital y respuesta de las plataformas. La política pública con perspectiva de género necesita un aterrizaje territorial.

La violencia digital no es menos real porque ocurra en una pantalla. La vergüenza, el miedo, la persecución, la pérdida de trabajo, el aislamiento social y el daño reputacional son materiales. La dimensión virtual sólo cambia la logística del agresor: abarata costos, amplifica el alcance, acelera la repetición y dificulta la reparación. En Ola Violeta decimos que, sin responsabilidad por parte de las plataformas, sin aplicación de la ley y sin perspectiva de género en el diseño y la supervisión de la IA, este desarrollo tecnológico, lejos de representar progreso, se convierte en una infraestructura de violencia.



Contraste internacional, brecha de protección

El análisis del panorama global muestra el otro componente del problema: la desigualdad en la protección legal y en la capacidad de cumplimiento. Aunque la violencia se parezca, la respuesta estatal y la obligación de las plataformas no son las mismas en todas partes. ONU Mujeres publicó el 18 de noviembre de 2025 un dato clave para dimensionar la brecha normativa: cerca de la mitad de las mujeres y niñas del mundo, alrededor de 1.8 mil millones, carecen de protección legal frente al abuso digital. Este dato encuadra el problema como un déficit de cobertura y de aplicación de normas, no como conductas aisladas de los usuarios.

En el mismo marco, ONU Mujeres resume rangos de prevalencia internacionales y reporta que estudios sitúan la violencia facilitada por la tecnología contra las mujeres en una afectación del 16% al 58%, según el país y la metodología. El rango amplio refleja diferencias en la medición y en la denuncia. En contraste, México ya cuenta con una medición anual oficial, el MOCIBA del INEGI, con desagregación por sexo y edad, lo cual facilita comparaciones consistentes dentro del país y pone en evidencia el reto de estandarizar métricas a nivel global. Cuando se trata de violencia sexual o sexualizada, el subregistro tiene causas conocidas: miedo, vergüenza, culpa inducida, dependencia económica, amenazas y desconfianza institucional. La tecnología no inventa ese silencio: lo explota.

La brecha legal no implica ausencia total de normas. También significa la ausencia de mecanismos de cumplimiento que obliguen a las plataformas a prevenir, detectar y retirar contenido, así como a conservar evidencia. En enero de 2026, en Gran Bretaña, Ofcom, la agencia gubernamental reguladora de los servicios de comunicación, abrió una investigación formal contra X bajo el Online Safety Act, con foco en los deberes de seguridad frente a contenido ilegal. En la



Unión Europea, el Digital Services Act se utiliza para exigir información y fomentar el cumplimiento cuando existen riesgos sistémicos. Esto contrasta con contextos en los que la norma existe, pero no hay supervisión técnica ni sanción efectiva.

El rango internacional de afectación de 16% a 58% que reporta ONU Mujeres no es un detalle estadístico menor. Muestra que cada país mide distinto, define distinto y enfrenta subregistro por miedo, vergüenza, dependencia económica o desconfianza institucional. Ese problema de medición impide comparaciones directas y oculta magnitudes. Por eso importan reportes con datos verificables de la plataforma, como el de AI Forensics sobre Grok, porque hacen visible lo que en otros contextos queda sin registro público. La evidencia técnica no sustituye la justicia, pero puede obligar a dar respuesta pública.

Infancias y material sexual ilegal, aceleración con IA

Cuando la tecnología vulnera a las infancias, no estamos ante malos usos aislados, sino ante un ecosistema criminal. Proteger a niñas y niños implica también entender que, históricamente, las niñas han sido el principal objetivo de la explotación sexual. Lamentablemente, lo digital hereda esa asimetría y la acelera. Internet Watch Foundation documentó en julio de 2025 un salto en la sofisticación y el volumen de material sobre abuso sexual infantil, en especial en formato de video. Esto confirma que la generación con IA ya no se limita a imágenes estáticas y que el riesgo aumenta cuando el contenido se vuelve más realista y fácil de producir y compartir.

Los medios que citan a la IWF reportaron un aumento del 400% en el número de páginas con material de abuso sexual infantil generado con IA en la primera



mitad de 2025, pasando de 42 a 210. Este dato no depende del caso Grok, pero sí marca una tendencia relevante. Cuando una plataforma integra la generación de imágenes sin controles eficaces, el ecosistema criminal ya cuenta con incentivos y redes para explotar esa capacidad a gran escala, con volumen y anonimato como ventajas operativas. La IWF también cuantificó el salto en video. Entre enero y junio de 2025 verificó 1,286 videos ilegales de abuso sexual infantil generados con IA, frente a 2 en el mismo periodo de 2024. Entre esos 1,286, la IWF clasificó 1,006 como Categoría A, el nivel más grave, lo que muestra que el crecimiento no se limita al contenido insinuado; también incluye material extremo.

En paralelo, la IWF reportó 210 páginas web con este material en la primera mitad de 2025, frente a 42 en 2024. Ese incremento del 400% indica que ya existe una infraestructura de publicación y redistribución que se adapta rápidamente a nuevas herramientas. En términos operativos, la IA reduce los costos de producción y acelera la rotación de contenidos. El núcleo aquí es de seguridad: si la oferta aumenta y el costo baja, la prevención no puede depender de reacciones lentas. Se necesitan barreras en origen, un monitoreo serio y cooperación con las autoridades, sin convertir a las víctimas en responsables de acreditar el daño. El hilo se conecta con Grok cuando la discusión pasa de la tendencia general a la evidencia de uso específico. En enero de 2026, *The Guardian* reportó que analistas de la IWF identificaron imágenes de abuso sexual infantil generadas con Grok, lo que refuerza la idea de que el riesgo no se limita a un mercado externo; también alcanza las herramientas integradas a las plataformas más usadas en el mundo.



Sesgos en modelos generativos

Los modelos aprenden del mundo tal como es —sesgado— y, si no se corrigen, lo devuelven como si las prácticas machistas fueran normales y no hubiese que superarlas. La neutralidad técnica es un mito cuando el entrenamiento se alimenta de la desigualdad. La UNESCO documentó que los modelos de lenguaje reproducen estereotipos de género cuando se entrenan con textos que ya reflejan esa desigualdad. En el ejemplo citado por la UNESCO, el modelo describió a mujeres en roles domésticos cuatro veces más que a los hombres. También asoció nombres femeninos con “hogar”, “familia” y “niños”, mientras que vinculó nombres masculinos con “negocio”, “ejecutivo”, “salario” y “carrera”. Este patrón muestra que el sesgo se manifiesta como una asociación automática entre el género y el rol social.

Ese sesgo nace de dos capas. La primera es el conjunto de datos. Si el corpus incluye más frases y narrativas que sitúan a mujeres en el cuidado y a hombres en el poder económico, el modelo aprende esa distribución y la devuelve en forma de texto plausible. La segunda capa es el diseño del sistema, las reglas de seguridad, la evaluación y el refuerzo con retroalimentación humana. Si el diseño no penaliza las salidas estereotipadas, el sistema las mantiene porque funcionan para completar patrones. El sesgo, sin embargo, es una posición política incorporada, aunque sea por omisión. Si el modelo asocia a las mujeres con el cuidado y a los hombres con la autoridad, está contribuyendo a un ecosistema simbólico en el que la desigualdad se percibe como natural. Y cuando ese ecosistema se conecta con plataformas y con la atención algorítmica, los estereotipos se vuelven infraestructura cultural.

Cuando el sesgo se combina con herramientas generativas que convierten texto en imagen, el daño cambia de nivel. Ya no se trata de una frase estereotipada;



se trata de una salida visual que sexualiza, degrada o expone. El caso Grok ilustra esta transición. El sesgo deja de ser un cliché textual y se convierte en una producción masiva de sexualización dirigida. También entra un componente etario que refuerza la asimetría, pues conecta con el estereotipo ideal de juventud y lo convierte en una norma visual repetida a escala.

El problema no se agota en una aplicación ni en un uso indebido. Si el sistema aprende asociaciones que colocan a las mujeres como objeto y a los hombres como agente, la salida tenderá a reforzar ese orden cuando el usuario lo solicite o cuando la plataforma premia la interacción con ese tipo de contenidos. En las plataformas sociales, la difusión, el archivo y el incentivo a la viralidad amplifican el sesgo porque transforman cada salida en contenido replicable.

Marco jurídico, más lento que la tecnología

Si el daño existe, debe haber herramientas para detenerlo, investigarlo y repararlo. La violencia digital no es un delito menor: es la continuidad de la violencia sexual y del control sobre la intimidad. La Ley Olimpia contiene un conjunto de reformas que reconocen la violencia digital como violencia de género y que tipifican conductas vinculadas a la intimidad sexual. Un punto clave del fenómeno de la IA es que la definición legal de violencia digital incorpora imágenes, audios o videos “reales o simulados” de contenido íntimo sexual difundidos sin consentimiento. Ese término “simulados” abre la puerta a contenidos sintéticos, incluidos “deepfakes” y “nudify”, cuando estos exponen o recrean la intimidad de una mujer sin su autorización.



En el plano penal federal, el Código Penal Federal incorporó el delito de “violación a la intimidad sexual” en el artículo 199 Octies. El tipo penal comprende dos bloques de conducta. Primero, la difusión no consentida por quien “divulgue, comparta, distribuya o publique” imágenes, videos o audios de contenido íntimo sexual de una persona mayor de edad, sin consentimiento, aprobación o autorización. Segundo, la producción o captación no consentida de quien “videograbe, audiograbe, fotografíe, imprima o elabore” imágenes, audios o videos con contenido sexual íntimo sin consentimiento, aprobación o autorización. La reforma no se quedó en el “contenido real”. El artículo 199 Nonies prevé las mismas sanciones cuando las imágenes, videos o audios difundidos “no correspondan con la persona que es señalada o identificada” en ellos. Esa redacción se alinea con el daño típico de los “deepfakes” sexuales: contenido atribuido a una persona que no participó en su producción.

Las penas base del 199 Octies fijan un umbral de gravedad. La reforma establece de tres a seis años de prisión y multa de 500 a 1,000 UMA. Además, el artículo 199 Decies incrementa el mínimo y el máximo hasta en una mitad cuando concurren agravantes que reflejan patrones frecuentes de violencia digital: relación sentimental o de confianza, servidor público en funciones, situación de vulnerabilidad de la víctima, obtención de un beneficio no lucrativo, fines de lucro, o cuando el impacto lleve a la víctima a atentar contra su integridad o su vida. Las agravantes reconocen un patrón que las víctimas han expresado durante años: que esta violencia suele venir de alguien cercano, de alguien con poder o de un entorno que explota vulnerabilidades.

El marco también incorpora una dimensión operativa de protección y de evidencia. La reforma prevé que la autoridad solicite el resguardo y la conservación del contenido denunciado y que la plataforma notifique al usuario que compartió el material que dicho contenido será inhabilitado por orden judicial. También fija



una audiencia dentro de cinco días para ratificar, modificar o cancelar las medidas de protección. El punto crítico, sin embargo, es que la eficacia para impedir estas agresiones contra las niñas y las mujeres no depende sólo del texto legal, sino también del acceso real a la justicia, de capacidades técnicas, de tiempos de respuesta y de la cooperación de las plataformas. En los casos de violencia digital, la demora forma parte del daño: cada hora sin retiro puede significar la multiplicación del alcance del material ofensivo.

El caso Grok funciona como una radiografía: muestra qué ocurre cuando una tecnología capaz de “crear” imágenes se integra a una plataforma de distribución masiva, con incentivos de viralidad y sin controles efectivos. Pero el problema no es Grok, el problema es el ecosistema: violencia digital ya extendida (y medida en México), una brecha global de protección (señalada por ONU Mujeres), una aceleración del riesgo extremo hacia las infancias (documentada por la IWF) y un trasfondo técnico y cultural en el que se replican estereotipos de género (documentado por UNESCO). Todo converge en que la sexualización sin consentimiento es violencia.

La conclusión feminista también es una exigencia pública. La carga de responsabilidad jurídica debe recaer en quien diseña, despliega, monetiza y gobierna las plataformas y los modelos. Eso implica controles preventivos, trazabilidad, retiro rápido, conservación de evidencia, sanciones efectivas y transparencia sobre las fallas y las mitigaciones. Si 2026 inicia con este escándalo, la pregunta es qué vamos a normalizar como sociedad. Porque cuando la violencia contra las niñas y las mujeres se vuelve tendencia, deja de ser un hecho aislado: se convierte en cultura. Y una cultura que tolera la producción masiva de material sexualizado sin consentimiento está diciendo, en voz alta, que el cuerpo de las mujeres es un recurso público. Ola Violeta sostiene lo contrario: la tecnología no



puede ser permiso para agredir. La innovación sin derechos no es progreso: es regresión. Y poco se puede esperar de Cupido en un contexto así.

“El mes del amor y la amistad es buen momento para cuestionarnos cómo la Inteligencia Artificial ha modificado la manera de iniciar y desarrollar relaciones sexoafectivas, en especial entre adolescentes y jóvenes”

Dra. María Elena Esparza Guevara

Fundadora de Ola Violeta AC

©Todos los derechos reservados Ola Violeta AC