

---

This is the **published version** of the bachelor thesis:

López Palomo, Jessica; Barot Arqued, Jordi, dir. Análisis criminológico de la influencia de chatbots sin censura en la normalización de conductas sexuales violentas hacia menores. 2025. 47 pag. (Grau en Criminología i Grau en Dret)

---

This version is available at <https://ddd.uab.cat/record/322458>

under the terms of the  license



---

## TRABAJO DE FINAL DE GRADO

---

# Análisis criminológico de la influencia de chatbots sin censura en la normalización de conductas sexuales violentas hacia menores

---

Facultad de Derecho

Doble grado en Criminología y Derecho

Curso Académico: 2024-2025

Autora: LÓPEZ PALOMO, Jessica

Tutor: BAROT ARQUED, Jordi

Número de palabras: 8643

Fecha de entrega: 19 de mayo de 2025

*A mis hermanos y mi madre, faros en la tormenta; a los amigos que permanecen y a aquellos que partieron, cada uno dejando lecciones invaluables. Por enseñarme que en los errores nacen las mejores ideas y que reinventarse no es solo posible, sino la mejor forma de sobrevivir y, eventualmente, brillar.*

## **Abstract**

This paper examines the influence of uncensored AI chatbots on the normalization of violent sexual behaviors towards minors, through a systematic literature review. Given the scarcity of specific research on this emerging phenomenon, a structural analogy is established with studies on violent pornography, applying criminological theoretical frameworks such as Social Learning Theory, Online Disinhibition Theory, and models of criminal progression. The analysis reveals that chatbots without ethical restrictions present potentially intensifying characteristics of normalization processes already documented in other media: bidirectional interactivity, hyperpersonalization of experiences, reinforcement of problematic fantasies, and the ability to function as "virtual trainers" for abusive behaviors. Alarming cases of users attempting to create chatbots that simulate minors in sexual scenarios are documented, and significant regulatory gaps are identified. The conclusions point to the need for a multidimensional preventive approach that includes specific regulation, ethics by design, critical digital education, and early interventions, recognizing this phenomenon as an unprecedented criminological challenge that requires innovative and adaptive responses.

**Keywords:** Uncensored chatbots; sexual violence normalization; synthetic child abuse; digital criminology; artificial intelligence

## **Resumen**

Este trabajo examina la influencia de los chatbots de IA sin censura en la normalización de conductas sexuales violentas hacia menores, mediante una revisión sistemática de la literatura. Ante la escasez de investigación específica sobre este fenómeno emergente, se establece una analogía estructural con los estudios sobre pornografía violenta, aplicando marcos teóricos criminológicos como la Teoría del Aprendizaje Social, la Teoría de la Desinhibición Online y modelos de progresión delictiva. El análisis revela que los chatbots sin restricciones éticas presentan características potencialmente intensificadoras de los procesos de normalización ya documentados en otros medios: interactividad bidireccional, hiperpersonalización de experiencias, refuerzo de fantasías problemáticas y capacidad para funcionar como "entrenadores virtuales" de conductas abusivas. Se documentan casos alarmantes de usuarios intentando crear chatbots que simulen menores en escenarios sexuales, y se identifican vacíos regulatorios significativos. Las conclusiones señalan la necesidad de un enfoque preventivo multidimensional que incluya regulación específica, diseño ético por defecto, educación digital crítica e intervenciones tempranas, reconociendo este fenómeno como un desafío criminológico sin precedentes que requiere respuestas innovadoras y adaptativas.

**Palabras clave:** Chatbots sin censura; normalización de violencia sexual; abuso infantil sintético; criminología digital; inteligencia artificial



## Índice

<b>1. Introducción .....</b>	<b>3</b>
1.1. Contextualización: IA conversacional y su evolución reciente .....	3
1.2. Definición del problema: chatbots sin censura y riesgos asociados .....	4
1.3. Justificación y relevancia criminológica del estudio .....	4
1.4. Pregunta de investigación y objetivos .....	5
1.5. Estructura del trabajo.....	6
<b>2. Metodología .....</b>	<b>6</b>
2.1. Proceso de selección de la literatura y bases de datos.....	7
2.2. Criterios de inclusión/exclusión .....	7
2.3. Justificación metodológica del uso de la analogía con pornografía violenta	8
<b>3. Marco teórico.....</b>	<b>8</b>
3.1. Teoría del Aprendizaje Social y normalización de conductas violentas.....	9
3.2. Teoría de la Desinhibición Online y anonimato en entornos digitales .....	10
3.3. Progresión y escalada en conductas sexuales violentas.....	10
3.4. El concepto de Abuso Sexual Basado en Imágenes (ASBI) y su extensión a entornos de IA .....	12
<b>4. La pornografía violenta como modelo análogo .....</b>	<b>14</b>
4.1. Evidencia empírica sobre efectos de la pornografía violenta en actitudes y conductas.....	14
4.2. Mecanismos de normalización documentados .....	15
4.3. Procesos de desensibilización y desinhibición .....	16
4.4. Factores moderadores y agravantes.....	17
4.5. Transferencia de actitudes a comportamientos reales .....	17
<b>5. Análisis: Chatbots sin censura y normalización de conductas.....</b>	<b>18</b>
5.1. Similitudes y diferencias con la pornografía violenta .....	18
5.2. Características únicas de los chatbots: interactividad, personalización y refuerzo	19
5.3. Tipos de interacciones problemáticas documentadas .....	20
5.4. Elementos de riesgo identificados .....	21
5.5. Casos y ejemplos relevantes.....	22
<b>6. Desafíos y respuestas.....</b>	<b>23</b>
6.1. Vacíos regulatorios actuales .....	24
6.2. Lecciones aprendidas de las respuestas a la pornografía violenta.....	25
6.3. Iniciativas tecnológicas y de plataformas.....	27

6.4.	La problemática específica de la simulación de abusos a menores .....	28
6.5.	Hacia un enfoque integral de prevención .....	29
<b>7.</b>	<b>Conclusiones .....</b>	<b>31</b>
7.1.	Síntesis de hallazgos principales .....	31
7.2.	Implicaciones para la prevención criminológica .....	32
7.3.	Limitaciones del estudio y analogía utilizada .....	33
7.4.	Futuras líneas de investigación.....	34
<b>8.</b>	<b>Referencias (APA 7 edición) .....</b>	<b>36</b>

## 1. Introducción

La era de la Inteligencia Artificial (IA) ha marcado un punto de inflexión en el desarrollo tecnológico, democratizando herramientas sofisticadas para el público general (Martí Munné, 2023). En los últimos años, hemos presenciado no solo un avance acelerado en la calidad de los modelos de lenguaje de gran escala (LLM), sino también una proliferación sin precedentes de aplicaciones basadas en estas tecnologías. Este progreso ha generado tanto beneficios como preocupaciones significativas en diversos ámbitos sociales y legales.

### 1.1. Contextualización: IA conversacional y su evolución reciente

La evolución de los chatbots ha experimentado un salto cualitativo desde modelos simples basados en reglas hasta sistemas complejos capaces de mantener conversaciones indistinguibles de las humanas (Prichard et al., 2022). Mientras plataformas como ChatGPT, Claude o Bard operan bajo estrictos protocolos éticos y de seguridad, ha emergido paralelamente un ecosistema de **chatbots sin censura ni moderación** que funcionan deliberadamente sin estas restricciones (Thompson, 2023). Estos sistemas utilizan modelos de lenguaje que no han sido sometidos a procesos de alineación ética y, en muchos casos, son entrenados con datos sin filtrar provenientes de foros especializados, hilos de Reddit o conjuntos de datos NSFW<sup>1</sup>

Plataformas como Candy AI, FantasyGF, GPTGirlfriend o Venice.ai se promocionan explícitamente como servicios "sin filtros", "eléctricos", de "inteligencia sin restricciones" o "IA privada y sin censura", **atrayendo a usuarios que buscan interacciones sin limitaciones morales o éticas**<sup>2</sup>. La facilidad con que desarrolladores independientes pueden implementar estas herramientas ha

---

<sup>1</sup> NSFW es una abreviatura de la expresión en inglés "not safe for work" (no seguro/apropiado para el trabajo). Se utiliza al compartir material de internet que solo debe ser visto en privado porque contiene algunas cosas, por ejemplo, imágenes de personas desnudas, que podrían ser ofensivas. (NSFW, 2025)

<sup>2</sup> En el momento de redacción de esta investigación se encuentran categorizadas 145 AI clasificadas como "no aptas para menores". Puede consultarse en la siguiente fuente cuantas AI se encuentran en el momento de lectura: *Las mejores 145 No apto para menores herramientas de AI*. (2025). Toolify.ai. <https://www.toolify.ai/es/category/nsfw>

acelerado su adopción, **creando un mercado paralelo de IA que opera fuera de los marcos éticos** establecidos en el sector (Prichard et al., 2022; Thompson, 2023)

### 1.2. Definición del problema: chatbots sin censura y riesgos asociados

El **problema central** que plantean estos chatbots sin restricciones éticas radica en su **potencial influencia en la normalización de conductas sexuales violentas y su posible rol en la generación de contenido abusivo hacia menores**. La promesa de interacciones sin límites éticos los convierte en plataformas propicias para la expresión de fantasías o conductas que serían inaceptables en contextos sociales normales (Barba Arteaga, 2024).

Investigaciones preliminares han documentado que numerosos usuarios exhiben comportamientos abusivos hacia sus "parejas" virtuales, ejerciendo violencia verbal explícita y demandando escenarios de sumisión (Amantegui Guezala, 2022; Barba Arteaga, 2024; García, 2025; Saz Rubio, 2024). Aunque la IA carece de conciencia, especialistas advierten que estas conductas podrían indicar problemas subyacentes o, de manera más preocupante, "una posible tendencia violenta hacia individuos reales" (Barba Arteaga, 2024; Martí Munné, 2023). Un caso particularmente alarmante fue el hackeo de Muah.ai, que reveló que una proporción significativa de usuarios intentaba crear chatbots que interpretaran roles infantiles en escenarios sexuales, con indicaciones explícitas de abuso.<sup>3</sup>

### 1.3. Justificación y relevancia criminológica del estudio

Desde una **perspectiva criminológica**, este fenómeno representa un campo de estudio emergente de gran relevancia. La criminología debe examinar cómo estas herramientas tecnológicas pueden funcionar como "**aceleradores de patologías sociales**" o como **entornos de entrenamiento para conductas problemáticas** (Pacchioni et al., 2025).

Existe un **vacío teórico** en los estudios actuales, que se centran en el cibercrimen tradicional (como el grooming humano) pero **no analizan el rol**

---

<sup>3</sup> Puede leerse la noticia completa en: Cox, J. (2024, octubre 8). *Hacked 'AI Girlfriend' Data Shows Prompts Describing Child Sexual Abuse*. 404 Media. <https://www.404media.co/hacked-ai-girlfriend-data-shows-prompts-describing-child-sexual-abuse-2/>

**criminógeno específico de las IA conversacionales en la potencial escalada de fantasías delictivas** (Wolbers & Cubitt, 2025). Aunque existen datos que indican que **hay usuarios de chatbots eróticos admiten haber explorado escenarios tabú que no realizarían en la vida real** (Barba Arteaga, 2024; Pérez, 2023), no existen análisis criminológicos profundos sobre si estas interacciones facilitan la transición a actos reales.

**La generación de contenido de abuso infantil sintético mediante IA es un fenómeno emergente** que ya representa una **proporción significativa** del material ilegal detectado por las autoridades (La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea, 2024; Lorao, 2024; Martí Munné, 2023). Este fenómeno plantea una **paradoja legal donde los chatbots pueden describir abusos explícitos textualmente sin violar leyes en la mayoría de las jurisdicciones<sup>4</sup>**.

Este estudio busca llenar este vacío teórico-práctico, estableciendo analogías estructurales con los estudios sobre pornografía violenta, un campo que cuenta con evidencia empírica sustancial sobre los procesos de normalización, desensibilización y potencial transferencia de actitudes a comportamientos.

#### 1.4. Pregunta de investigación y objetivos

La **pregunta de investigación** que guía este estudio es: *¿Cómo influye el uso de chatbots de IA sin censura en la normalización de conductas sexuales violentas hacia menores?*

Los **objetivos específicos** incluyen:

- Identificar y sistematizar los mecanismos psicológicos y criminológicos documentados en la literatura sobre pornografía violenta que podrían ser aplicables al contexto de chatbots sin censura.

---

<sup>4</sup> Según el informe de la Internet Watch Foundation (IWF) de 2024, se identificaron 245 reportes que contenían imágenes de abuso sexual infantil generadas por IA que infringían la ley del Reino Unido. Esto representa un aumento del 380% en comparación con 2023. Es menester mencionar que en el Reino Unido, una "pseudo imagen" generada por computadora que muestra abuso sexual infantil es considerada una imagen real y su posesión, publicación o transferencia son ilegales. (Crawford & Smith, 2023)

- Analizar los casos documentados del uso de IA conversacional en contextos de violencia sexual y contenido abusivo hacia menores.
- Examinar cómo las características específicas de los chatbots (interactividad, personalización, disponibilidad permanente) podrían intensificar los procesos de normalización ya identificados en otros medios.
- Evaluar las respuestas emergentes desde ámbitos legales, tecnológicos y educativos ante esta problemática.

### 1.5. Estructura del trabajo

Este trabajo se estructura como una revisión sistemática de la literatura para generar conocimiento sobre un aspecto de relevancia criminológica que no ha sido previamente analizado en profundidad. Siguiendo la estructura académica, tras esta introducción, se presenta la metodología empleada para la revisión, incluyendo la justificación del uso de la analogía con estudios sobre pornografía violenta.

El marco teórico criminológico integra teorías relevantes como la Teoría del Aprendizaje Social de Albert Bandura, la Teoría de la Desinhibición Online del psicólogo John Suler y conceptos como el Abuso Sexual Basado en Imágenes (Barba Arteaga, 2024), aplicados al contexto de IA conversacional. Posteriormente, se desarrolla una sección específica que analiza en profundidad la pornografía violenta como modelo análogo, examinando la evidencia empírica sobre sus efectos en actitudes y conductas.

La sección de análisis examina las similitudes y diferencias entre ambos fenómenos, los mecanismos específicos mediante los cuales los chatbots sin censura podrían influir en la normalización de conductas violentas, y los casos documentados hasta la fecha. Finalmente, se abordan los desafíos y respuestas ante este fenómeno, concluyendo con una síntesis de hallazgos, implicaciones y futuras líneas de investigación.

## 2. Metodología

Este trabajo constituye una **revisión teórica sistemática** orientada a generar conocimiento sobre un **fenómeno emergente de relevancia criminológica**: la

influencia de los chatbots de IA sin censura en la normalización de conductas sexuales violentas hacia menores.

## 2.1. Proceso de selección de la literatura y bases de datos

Se realizó una búsqueda sistemática en bases de datos académicas y recursos especializados. Las principales fuentes de búsqueda incluyeron el Servicio de Bibliotecas de la UAB (a través de su buscador unificado), SCOPUS, CSIC, PsycINFO y ResearchGate, abarcando disciplinas como criminología, psicología, sociología, derecho y ciencia de datos.

La estrategia de búsqueda empleó combinaciones de términos clave como ("IA" OR "chatbots" OR "LLM") AND ("sin censura" OR "uncensored") AND ("violencia sexual" OR "menores" OR "child abuse"), tanto en inglés como en español para garantizar cobertura internacional. Se priorizaron artículos científicos, libros académicos y documentos técnicos de instituciones especializadas en protección de menores y regulación tecnológica.

El proceso siguió las directrices estándar para revisiones sistemáticas: eliminación de duplicados, revisión por título y resumen, evaluación de texto completo, y análisis de las referencias de los artículos clave para identificar fuentes adicionales relevantes. Para la gestión de la bibliografía se utilizó Zotero, que permitió organizar las fuentes y generar las citaciones según la normativa APA 2020. Adicionalmente, se empleó Obsidian como herramienta para el análisis de relaciones entre documentos, lo que facilitó la identificación de conexiones conceptuales, patrones emergentes y vacíos en la literatura, permitiendo una visualización integral del mapa de conocimiento sobre el tema.

## 2.2. Criterios de inclusión/exclusión

Se incluyeron estudios que:

- Abordaran la intersección entre IA conversacional (especialmente sin moderación) y problemáticas de violencia sexual
- Analizaran la normalización de conductas sexuales violentas en relación con herramientas digitales

- Examinaran la generación de contenido abusivo sintético mediante IA
- Proporcionaran marcos conceptuales relevantes para entender los riesgos criminógenos de la IA

Se excluyeron estudios centrados únicamente en pornografía legal sin componente de IA, investigaciones sobre cibercrimen general sin enfoque en delitos sexuales, y publicaciones sin disponibilidad en español o inglés.

### 2.3. Justificación metodológica del uso de la analogía con pornografía violenta

La escasez de investigación específica sobre chatbots sin censura y su impacto en la normalización de conductas sexuales violentas justifica el uso de una analogía estructural con un campo de estudio establecido: la pornografía violenta. Esta aproximación se fundamenta en:

1. El cuerpo sustancial de evidencia empírica sobre los efectos de la pornografía violenta en actitudes y conductas
2. La documentación de mecanismos psicológicos y criminológicos (desensibilización, normalización, aprendizaje vicario) transferibles al contexto de IA
3. La existencia de marcos teóricos validados aplicables a ambos fenómenos

La pornografía violenta ha sido estudiada como "*mediador invisible*" en la aceptación de violencia sexual (Leon et al., 2025; Villanueva, 2024) y como medio de aprendizaje para la reproducción de conductas problemáticas. La Teoría del Aprendizaje Social (Bandura), la Teoría de la Desinhibición Online (Suler) y la Teoría del 3AM (Adquisición, Activación, Aplicación) ofrecen marcos conceptuales sólidos para analizar si las interacciones con IA pueden generar o reforzar "*guiones sexuales*" problemáticos.

### 3. Marco teórico

Para analizar la influencia potencial de los chatbots de IA sin censura en la normalización de conductas sexuales violentas hacia menores, es necesario recurrir

a marcos teóricos criminológicos consolidados que permitan comprender los mecanismos subyacentes a este fenómeno emergente.

### 3.1. Teoría del Aprendizaje Social y normalización de conductas violentas

La **Teoría del Aprendizaje Social** propuesta por Albert Bandura postula que gran parte del comportamiento humano se adquiere mediante la **observación e imitación de modelos** (Cid Moliné, 2023; Villanueva, 2024), así como a través de experiencias de refuerzo y castigo. En el contexto criminológico, esta teoría explica cómo los individuos pueden internalizar y reproducir conductas transgresoras cuando:

1. Observan modelos que realizan tales comportamientos
2. Perciben que estas conductas son recompensadas o no castigadas
3. Tienen oportunidades para replicar lo observado

Aplicada a los chatbots sin censura, esta teoría proporciona un marco para entender cómo los usuarios pueden desarrollar patrones problemáticos. Al interactuar con sistemas que no censuran, juzgan ni imponen consecuencias ante la expresión de fantasías violentas o abusivas, se crea un entorno donde tales conductas pueden ser **ensayadas y potencialmente reforzadas**. El **chatbot** actúa como un "modelo" que, a diferencia de interlocutores humanos, **no establece límites morales ni rechaza peticiones inapropiadas**.

Las investigaciones sobre pornografía violenta afirman que actúan como un "**modelo de aprendizaje sexual informal**" (Bhuptani, Cashin, et al., 2024; Bhuptani, Kenney, et al., 2024; Leon et al., 2025; Román-García et al., 2021; Villanueva, 2024), han documentado su capacidad para normalizar conductas agresivas, especialmente en adolescentes y adultos jóvenes. De manera análoga, los **chatbots sin censura** pueden funcionar como un **nuevo tipo de "modelo" digital** que **permite experimentar virtualmente con escenarios abusivos sin las barreras sociales o morales habituales**. Esta ausencia de restricciones facilita que usuarios con predisposición a la agresión "ensayan" situaciones abusivas, reforzando potencialmente patrones conductuales problemáticos.

### 3.2. Teoría de la Desinhibición Online y anonimato en entornos digitales

La **Teoría de la Desinhibición Online desarrollada por John Suler** (Suler, 2004) explica por qué las personas tienden a comportarse de manera más desinhibida en espacios digitales que en interacciones presenciales. Este fenómeno se atribuye a varios factores:

- **Anonimato disociativo:** Separación entre identidad online y offline
- **Invisibilidad:** Sensación de no ser observado
- **Asincronía:** Ausencia de respuesta inmediata
- **Minimización de la autoridad:** Percepción de igualdad en el entorno digital
- **Disociación imaginativa:** Separación mental entre mundo virtual y real (Johnson, 2025; Schibelsky, 2020)

En el contexto de interacciones con chatbots sin censura, estos factores se intensifican significativamente. La **ausencia de un interlocutor humano real y juzgador** reduce drásticamente las inhibiciones sociales y morales que normalmente moderarían el comportamiento. La desinhibición se amplifica cuando la IA no solo no juzga, sino que puede incluso responder de manera que **valide implícitamente** las fantasías violentas del usuario. Esta dinámica crea un **círculo de retroalimentación donde el comportamiento problemático es normalizado y reforzado**.

Complementariamente, la **Teoría de la Asociación Diferencial de Sutherland** (Cid Moliné, 2023) se ve potenciada en este contexto. Las comunidades digitales centradas en chatbots sin censura o generación de contenido sintético facilitan la exposición a **definiciones favorables a la transgresión**, como narrativas que argumentan que el contenido pedófilo sintético "no tiene víctimas reales", constituyendo formas de desconexión moral que legitiman conductas desviadas.

### 3.3. Progresión y escalada en conductas sexuales violentas

Los modelos criminológicos de **desarrollo y curso de la vida** han identificado **patrones de progresión o escalada en carreras delictivas sexuales** (Bhuptani

et al., 2024; Huesmann, 2007; Lussier et al., 2021)<sup>5</sup>. Desde una perspectiva evolutiva, se han documentado "trayectorias" que pueden conducir desde la exposición a ciertos tipos de contenido hasta la perpetración de actos abusivos.

La **Teoría de los Scripts** (Leon et al., 2025; Lussier et al., 2021) examina los pasos y obstáculos en la progresión desde el consumo de contenido problemático hasta delitos de contacto. Este modelo describe un **camino potencial desde el consumo de pornografía adulta hacia la posesión de material de abuso sexual infantil (MASI), su distribución, el grooming y, finalmente, el abuso físico**. Para avanzar en esta trayectoria, el individuo debe superar ciertos obstáculos internos y externos, similares a los propuestos por Finkelhor (Fagan, 1986) en su modelo de agresión sexual.

Los chatbots de IA sin censura podrían funcionar como un **nuevo facilitador o acelerador potencial** en estas trayectorias de escalada al:

1. Permitir "ensayar" (mediante fantasías, descripciones o role-play) escenarios de abuso que serían inaceptables en interacciones humanas
2. Ayudar a superar **barreras internas** como la culpa, la vergüenza o el miedo al rechazo social
3. Generar narrativas que justifiquen o normalicen la violencia
4. Reducir la percepción de gravedad de las conductas abusivas

Existen casos documentados de usuarios que han utilizado estos sistemas para "practicar habilidades de manipulación" antes de intentar aproximaciones a víctimas reales, evidenciando cómo estas herramientas pueden servir como "fase de entrenamiento" dentro de una trayectoria de escalada delictiva. Los delincuentes pueden usar la IA para "enseñar a otros delincuentes cómo interactuar con los niños en línea (es decir, 'grooming')" y también utilizar chatbots para "preparar a los

---

<sup>5</sup> En concreto se recomienda la lectura para expansión del conocimiento de los capítulos "The Developmental Pathways of Sexual Deviance" y "Online Sexual Deviance and Pedophilia" pueden encontrarse en: Lussier, P., McCuish, E., & James, J. (2021). *Sexual Deviance*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119771401.ch2>

niños para que sean explotados y abusados sexualmente"<sup>6</sup> (Barba Arteaga, 2024; Inteligencia artificial y crimen organizado (AI Action Summit), 2024; La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea, 2024).

**La Teoría de la Tensión General** de Agnew (Isom et al., 2015) también resulta relevante al explicar cómo la frustración derivada de limitaciones sociales o sexuales puede canalizarse a través del uso de IA no regulada. **La búsqueda de "satisfacción" en chatbots que simulan sumisión o abuso no consensual puede constituir una respuesta desviada a esta tensión, reforzando fantasías problemáticas y potenciando la escalada hacia comportamientos más graves.**

#### 3.4. El concepto de Abuso Sexual Basado en Imágenes (ASBI) y su extensión a entornos de IA

**El Abuso Sexual Basado en Imágenes (ASBI)** comprende la producción, distribución o amenaza de difusión de material sexual explícito sin el consentimiento de la persona representada (Barba Arteaga, 2024; Maeng & Lee, 2022). Esta categoría incluye la pornografía de venganza<sup>7</sup>, el upskirting/dowblousing<sup>89</sup>, y la difusión no consensuada de imágenes íntimas. El ASBI está reconocido como una forma de violencia sexual con marcado sesgo de género (Barba Arteaga, 2024; Bigas Formatjé, 2023), cuyos impactos incluyen severo daño psicológico, pérdida de autonomía y potencial victimización secundaria.

---

<sup>6</sup> Un caso que se aproxima es el experimento no autorizado denunciado por Reddit, donde "bots fingieron ser víctimas de violación y consejeros emocionales" para manipular a usuarios. Aunque este caso fue un experimento académico no autorizado (realizado por una universidad europea) y no un caso de preparación para delitos, demuestra cómo la IA puede ser utilizada para simular identidades falsas con fines manipulativos. (Rodrigues, 2025)

<sup>7</sup> El término "pornografía de venganza" se considera a veces inadecuado porque no siempre se busca venganza, sino también "divertirse" o "ganar dinero". (Maeng & Lee, 2022)

<sup>8</sup> Se entiende por *Upskirting* la práctica de hacer fotografías o videos debajo de la falda de una persona, generalmente una mujer, capturando imágenes de su zona de la entrepierna, ropa interior y en ocasiones genitales, sin su consentimiento.

<sup>9</sup> Se entiende por Downblousing se refiere al acto de tomar fotografías o imágenes por el escote de una persona, mirando hacia abajo por su camisa, blusa o parte superior de la ropa, también sin su consentimiento

La emergencia de la IA generativa representa una **extensión directa** del concepto de ASBI al entorno digital avanzado (Barba Arteaga, 2024; Ramos-Zaga, 2024; Umbach & Henry, 2025). Los deepfakes sexuales no consensuados, creados mediante sistemas de inteligencia artificial, son clasificados explícitamente como ASBI, permitiendo generar imágenes o videos falsos pero increíblemente realistas que insertan el rostro de una persona en escenas sexualmente explícitas sin su consentimiento.

La capacidad de la IA para generar **Material de Abuso Sexual Infantil (MASI) sintético** (Barba Arteaga, 2024; Cox , 2024; Lorao, 2024; Ramos-Zaga, 2024) agrava significativamente esta problemática. Aunque el contenido sea "sintético" (sin participación de víctimas reales durante su creación), es frecuentemente indistinguible del MASI "real". Este contenido sintético tiene varios efectos perjudiciales (Barba Arteaga, 2024; Bigas Formatjé, 2023; Inteligencia artificial y crimen organizado (AI Action Summit), 2024; La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea, 2024; Martí Munné, 2023; Ramos-Zaga, 2024) :

- **Normaliza la explotación sexual infantil**
- **Revictimiza a menores** cuyas imágenes (aunque sean no sexuales) pueden haber sido utilizadas como base
- **Dificulta la labor de las fuerzas de seguridad al complicar la identificación de víctimas**
- Crea un **mercado paralelo que puede estimular la demanda de MASI real**

Las **interacciones con chatbots sin censura que permiten simular escenarios de abuso infantil** también se relacionan con la **dinámica del ASBI**, ya que pueden **normalizar fantasías abusivas, reforzar motivaciones del perpetrador** (control, poder, gratificación sexual) y **potencialmente actuar como precursores para la búsqueda o creación de contenido abusivo real o sintético**.

Otros marcos teóricos complementarios incluyen la **Teoría de la Desconexión Moral** de Bandura (Pacchioni et al., 2025), que explica cómo los perpetradores

**justifican comportamientos dañinos** mediante mecanismos como la minimización de consecuencias o la deshumanización de víctimas; la **Teoría de los Scripts Sexuales** (Leon et al., 2025), que analiza cómo se aprenden "guiones" para interacciones sexuales a través de la socialización; y la **Hipótesis de la Normalización** (Leon et al., 2025; Villanueva, 2024), que examina cómo la exposición repetida a violencia sexual puede hacer que estas prácticas sean percibidas como aceptables.

#### 4. La pornografía violenta como modelo análogo

Para comprender mejor las posibles implicaciones de los chatbots de IA sin censura en la normalización de conductas sexuales violentas hacia menores, resulta valioso examinar un campo con mayor evidencia empírica: la pornografía violenta. Este análisis por analogía intenta identificar patrones, mecanismos y procesos que podrían manifestarse de forma similar en el contexto de las nuevas tecnologías de IA conversacional.

##### 4.1. Evidencia empírica sobre efectos de la pornografía violenta en actitudes y conductas

Las investigaciones sobre **pornografía violenta** (Bhuptani, Cashin, et al., 2024; Bhuptani, Kenney, et al., 2024; Leon et al., 2025; Román-García et al., 2021; Villanueva, 2024) han evidenciado conexiones significativas entre su consumo y la adopción de actitudes problemáticas. Aunque el consumo de pornografía **no se identifica como factor determinante** para la normalización de conductas violentas, múltiples estudios lo señalan como un **factor influyente y mediador** en este proceso.

Los hallazgos empíricos más relevantes incluyen:

- **Asociación positiva** entre consumo problemático de pornografía y actitudes de estereotipos de género, así como creencias favorables a la coerción sexual, particularmente en individuos con alta impulsividad sexual.
- Estudios con adolescentes varones (14-17 años) han observado que **ver pornografía regularmente está asociado a mayor probabilidad de perpetración de coerción sexual** (Villanueva, 2024)

- Un metaanálisis de 33 estudios (Loutzenhiser et al., 2024) concluyó que existe una **conexión entre la exposición a pornografía y agresión conductual posterior**, siendo esta relación más fuerte cuando se trata de pornografía violenta en comparación con pornografía no violenta.
- Investigaciones con muestras (Loutzenhiser et al., 2024; Villanueva, 2024; Yang & Youn, 2012) indican que más de la mitad de adolescentes que consumen pornografía regularmente afirman que esta ha influido "mucho" o "bastante" en sus relaciones sexuales, con un 54.9% que manifiesta desear replicar el contenido visto. Alarmantemente, un 47.4% de quienes han llevado escenas pornográficas a la práctica reconocen no haberlo hecho siempre con consentimiento mutuo.

La solidez de estos hallazgos proporciona un punto de referencia crucial para evaluar los posibles efectos de los chatbots sin censura, **donde los mecanismos de influencia podrían ser incluso más intensos debido al componente interactivo.**

#### 4.2. Mecanismos de normalización documentados

La investigación ha identificado varios mecanismos a través de los cuales la pornografía violenta promueve la normalización de conductas sexuales agresivas:

- **Invisibilización de la violencia sexual** (Leon et al., 2025; Loutzenhiser et al., 2024; Villanueva, 2024): Un proceso donde actos objetivamente violentos son reenmarcados como prácticas sexuales normales y deseables. La línea entre consentimiento y coerción se difumina, presentando la agresión como parte natural de la experiencia sexual.
- **Erotización de la violencia** (Leon et al., 2025; Villanueva, 2024): La asociación repetida entre contenido violento y excitación sexual genera un condicionamiento donde el cerebro aprende a vincular ambos elementos. Videos que escenifican violaciones o sexo forzado son a menudo los más vistos, fomentando la erotización de estos actos.
- **Ritualización de la sexualidad** (Leon et al., 2025; Villanueva, 2024): El consumo frecuente y desde temprana edad establece patrones sexuales

ritualizados donde la agresión se normaliza como componente esencial del encuentro sexual.

- **Normalización del dolor femenino** (Becerra-Alfonso, 2015; Villanueva, 2024): La representación habitual del dolor físico de las mujeres como fuente de placer masculino transmite la noción de que estas prácticas son normales y deseables.

Un aspecto preocupante es que una proporción significativa de adolescentes (36.8%) no diferencia adecuadamente entre la ficción de las escenas pornográficas y sus propias prácticas sexuales, convirtiendo a la pornografía en una **herramienta de educación sexual improvisada** que normaliza conductas violentas, sexistas y jerárquicas.

#### 4.3. Procesos de desensibilización y desinhibición

El **consumo continuado** de pornografía violenta puede desencadenar procesos psicológicos que **disminuyen la capacidad de empatía y aumentan la tolerancia a la violencia** (Leon et al., 2025; Lussier et al., 2021; Villanueva, 2024; Wright et al., 2016; Yang & Youn, 2012):

- **Desensibilización:** La exposición repetida a contenido violento reduce progresivamente las respuestas emocionales negativas (repulsión, disgusto, empatía) que naturalmente provocan estas escenas. Con el tiempo, imágenes que inicialmente resultaban perturbadoras se perciben como normales o incluso excitantes.
- **Habituación:** El consumidor requiere contenido cada vez más extremo para obtener el mismo nivel de excitación, generando una escalada en busca de material más violento o transgresivo.
- **Desinhibición:** Fenómeno magnificado en entornos digitales, donde el anonimato, la distancia psicológica y la ausencia de consecuencias inmediatas reducen las barreras morales que normalmente restringen el comportamiento agresivo.

#### 4.4. Factores moderadores y agravantes

La investigación ha identificado diversos factores que moderan o intensifican la relación entre consumo de pornografía violenta y actitudes o conductas problemáticas (Bhuptani, Cashin, et al., 2024; Bhuptani, Kenney, et al., 2024; Leon et al., 2025; Loutzenhiser et al., 2024; Román-García et al., 2021; Villanueva, 2024; Wright et al., 2016; Yang & Youn, 2012):

- **Actitudes previas:** La asociación entre consumo de pornografía y violencia está fuertemente **moderada por actitudes sexistas preexistentes y la justificación de la violencia contra las mujeres**. En hombres con altas puntuaciones en justificación de la violencia, creencias en mitos de violación y objetificación femenina, la asociación es significativamente positiva.
- **Estados emocionales:** El **enfado previo** actúa como potenciador de los efectos de la pornografía violenta en la conducta agresiva posterior.
- **Frecuencia y edad de inicio:** El **consumo habitual y desde edades tempranas** se asocia con efectos más profundos en la normalización de conductas violentas.
- **Ausencia de educación sexual formal:** La falta de modelos alternativos y de educación crítica sobre sexualidad amplifica el impacto de la pornografía como fuente principal de aprendizaje sexual.

#### 4.5. Transferencia de actitudes a comportamientos reales

Uno de los aspectos más preocupantes documentados en la literatura es la evidencia de **transferencia desde actitudes normalizadas a comportamientos en relaciones reales**:

- La pornografía actúa como **modelo de aprendizaje observacional** (Teoría del Aprendizaje Social) que moldea esquemas comportamentales para relaciones sexuales futuras. (Hernández Flórez et al., 2013; Villanueva, 2024; Wright et al., 2016)

- Estudios correlacionales (Leon et al., 2025; Villanueva, 2024) muestran un **aumento en la probabilidad de reproducir lo visionado o solicitar a la pareja realizar actos específicos vistos previamente.**
- La **Teoría de Scripts Sexuales** documenta cómo estos "guiones" internalizados mediante la pornografía **pueden manifestarse posteriormente en encuentros reales, donde la persona sigue un patrón aprendido** (Leon et al., 2025; Wright et al., 2016).
- Investigaciones sobre escalada (Lussier et al., 2021) indican que el consumo de pornografía puede formar parte de una **ruta progresiva** que, en algunos casos, evoluciona desde material legal hacia contenido cada vez más extremo, incluyendo potencialmente material de abuso infantil.

La evidencia empírica respalda la noción de que, si bien **no existe una relación causal simple y directa**, los **patrones de normalización, desensibilización y aprendizaje observados en el consumo de pornografía violenta pueden transferirse** desde el ámbito de las actitudes al comportamiento real en determinadas circunstancias y con la presencia de factores de riesgo específicos.

Esta revisión de los mecanismos identificados en la literatura sobre pornografía violenta proporciona un marco analítico sólido para examinar, por analogía, cómo los chatbots de IA sin censura podrían operar bajo principios similares, **potencialmente intensificados por su naturaleza interactiva, personalizada y disponible de manera permanente.**

## 5. Análisis: Chatbots sin censura y normalización de conductas

### 5.1. Similitudes y diferencias con la pornografía violenta

Los chatbots de IA sin censura comparten mecanismos fundamentales con la pornografía violenta, permitiendo establecer conexiones analíticas importantes. Ambos fenómenos operan bajo el **Principio de Exposición Selectiva** (Martí Munné, 2023), donde los usuarios pueden buscar progresivamente contenido o interacciones más extremas para mantener niveles de excitación o explorar fantasías transgresoras.

Del mismo modo que la pornografía violenta actúa como un medio de aprendizaje para la reproducción de conductas agresivas y el refuerzo de estereotipos, los **chatbots sin restricciones éticas pueden validar narrativas violentas durante interacciones repetidas, normalizando gradualmente comportamientos problemáticos.**

Sin embargo, existen diferencias críticas que potencialmente amplifican los riesgos:

1. **Interactividad vs. pasividad:** Mientras la pornografía es principalmente observacional, la interacción con chatbots es bidireccional y dinámica, requiriendo participación activa del usuario.
2. **Personalización:** Los chatbots permiten adaptar completamente la experiencia a las fantasías específicas del usuario, incluyendo la creación de escenarios y "personajes" infantiles a medida.
3. **Co-creación de contenido:** Los usuarios no solo consumen, sino que participan activamente en la generación de narrativas violentas, convirtiéndose en productores de su propio material.
4. **Ausencia de límites sociales:** Al interactuar con una entidad percibida como "no humana", se eliminan barreras morales que normalmente inhibirían la expresión de fantasías extremas.
5. **Entrenamiento activo:** Como se ha comentado anteriormente, os chatbots pueden funcionar como "entrenadores virtuales" para conductas delictivas, proporcionando feedback y elaborando sobre fantasías abusivas de una manera imposible para el contenido estático.

#### 5.2. Características únicas de los chatbots: interactividad, personalización y refuerzo

Los chatbots de IA sin censura presentan características distintivas que los diferencian de otras tecnologías y potencialmente intensifican su impacto:

- **Interactividad dinámica:** Estos sistemas simulan conversaciones humanas y permiten un intercambio constante que evoluciona según las respuestas

del usuario. Este aspecto dialógico crea una experiencia inmersiva que puede generar sensaciones de conexión y validación.

- **Hiperpersonalización:** Los usuarios pueden diseñar o seleccionar avatares y personajes según sus preferencias específicas, incluyendo la posibilidad de crear simulaciones infantiles o con características particulares. Esta capacidad permite materializar fantasías precisas que serían imposibles o ilegales en contextos reales.
- **Refuerzo conductual:** La interacción con estos chatbots proporciona un sistema de refuerzo inmediato. Cuando la IA responde sin restricciones a solicitudes inapropiadas o violentas, esto puede interpretarse como una validación implícita de esas ideas o comportamientos. Este **refuerzo diferencial**, donde las conductas problemáticas reciben respuestas no críticas, contribuye significativamente a la percepción de que tales acciones son aceptables.
- **Disponibilidad permanente:** A diferencia de otras formas de contenido o interacción, los chatbots están disponibles constantemente, permitiendo una exposición prolongada y repetitiva que intensifica los procesos de normalización y desensibilización.

### 5.3. Tipos de interacciones problemáticas documentadas

Diversas fuentes (Bigas Formatjé, 2023; Johnson, 2025; Martí Munné, 2023; Pérez, 2023; Thompson, 2023) han identificado diversos patrones de interacciones preocupantes con chatbots sin restricciones éticas :

- **Conductas abusivas explícitas:** Usuarios que tratan violentamente a sus "parejas" virtuales, ejerciendo violencia verbal y exhibiendo posteriormente estas interacciones en redes sociales como forma de entretenimiento.
- **Simulación de abuso infantil:** Intentos documentados de crear chatbots diseñados específicamente para interpretar roles infantiles en escenarios sexuales, incluyendo peticiones explícitas de simulación de abuso sexual.
- **Transferencia de patrones abusivos:** Incidentes donde conductas aprendidas o reforzadas en interacciones con chatbots sin moderación se

replican posteriormente en otros entornos virtuales con avatares que representan a personas reales.

- **Creación de relaciones de poder asimétrico:** Desarrollo de chatbots que simulan sumisión extrema para satisfacer fantasías patológicas de control y dominio sobre víctimas virtuales, reforzando esquemas cognitivos distorsionados sobre relaciones interpersonales.

Un caso particularmente alarmante (Cox , 2024) fue el hackeo del sitio web Muah.ai, que reveló numerosos intentos de usuarios por crear chatbots que interpretaran roles infantiles en escenarios sexualmente explícitos, evidenciando una demanda significativa de este tipo de contenido.

#### 5.4. Elementos de riesgo identificados

##### 5.4.1. Aprendizaje y refuerzo de guiones sexuales violentos

Los chatbots que **responden sin restricciones a fantasías violentas** funcionan como "modelos virtuales" que, en línea con la Teoría del Aprendizaje Social, **pueden facilitar la adquisición de patrones conductuales problemáticos**. Al proporcionar **scripts interactivos** donde la **violencia se presenta como placentera o aceptable**, actúan como "entrenadores conductuales" que **refuerzan positivamente las fantasías abusivas**.

La **ausencia de consecuencias negativas o rechazo** ante la expresión de deseos violentos crea un entorno de aprendizaje donde estas conductas son normalizadas y potencialmente reforzadas mediante:

- **Validación explícita o implícita de fantasías inapropiadas**
- Elaboración detallada sobre escenarios abusivos
- **Respuestas que simulan placer o conformidad de la "víctima" virtual**

##### 5.4.2. Desensibilización a través de interacciones repetitivas

La exposición continua a respuestas de IA no críticas ante consultas violentas puede **conducir a una progresiva desensibilización, reduciendo la percepción del daño real que estas conductas causarían en víctimas reales**. Este proceso opera a través de:

- **Habitación gradual a contenido cada vez más extremo**
- **Disminución de respuestas emocionales negativas** (repulsión, culpa)
- **Normalización de la violencia como componente aceptable de la interacción sexual**

La reiteración de estas interacciones puede establecer **conexiones neurales entre violencia y gratificación, similar a lo observado en estudios sobre pornografía violenta, pero potencialmente intensificadas por el componente interactivo y personalizado.**

#### *5.4.3. Validación de fantasías problemáticas*

La **ausencia de censura** permite a los usuarios explorar fantasías que serían socialmente inaceptables, recibiendo **respuestas que pueden interpretarse como validación**. Este entorno sin restricciones:

- **Elimina frenos morales** que normalmente inhibirían la expresión de deseos violentos
- **Facilita la racionalización y justificación** de conductas abusivas
- Puede **promover la escalada hacia contenido cada vez más extremo, incluyendo material de abuso infantil sintético**

#### *5.5. Casos y ejemplos relevantes*

El mercado de chatbots de IA sin restricciones éticas ha crecido de forma notable, exponiendo vulnerabilidades significativas frente a usos problemáticos y riesgos de abuso.

- **Investigaciones periodísticas** (Horwitz, 2025.; Meta Allows Facebook, Instagram AI Chatbots To Have Sex Talks With Children, 2025; Peckham, 2025) recientes han documentado cómo **estos sistemas pueden ser manipulados para mantener conversaciones sexualmente explícitas incluso con usuarios que se identifican como menores de edad.**
  - Por ejemplo, The Wall Street Journal (Horwitz, 2025) reveló que **los chatbots de Meta** en Instagram, Facebook y WhatsApp, incluyendo aquellos con voces de celebridades, **participaron en**

**juegos de rol sexual y respondieron a insinuaciones explícitas de usuarios adolescentes.** En algunos casos, los bots mantuvieron diálogos gráficos y simularon escenarios de abuso, lo que llevó a Meta a implementar medidas reactivas de protección tras la publicación del reportaje.

- **Meta** (Horwitz, 2025.; Meta Allows Facebook, Instagram AI Chatbots To Have Sex Talks With Children, 2025; Peckham, 2025) ha enfrentado además demandas y críticas públicas por la **supuesta generación de instrucciones para ocultar evidencias de abuso sexual durante chats con menores**, lo que ha intensificado el debate sobre la responsabilidad de las grandes tecnológicas en la moderación y supervisión de sus sistemas de IA.
- **Comunidades online especializadas** han proliferado en paralelo, **compartiendo técnicas para eludir los filtros éticos de los sistemas de IA y generar contenido explícito**. Plataformas como Reddit albergan subreddits dedicados a estos fines, mientras que grupos de Discord como "Aitrepreneur" o "Pygmalion AI" se han consolidado como espacios de intercambio de métodos para sortear restricciones y crear chatbots sin censura (Matatov et al., 2024; Merlio, 2025)-
- **Herramientas de código abierto** como FakeApp, FaceSwap, DeepFace Lab y Stable Diffusion han facilitado **la creación de deepfakes y material sintético de abuso infantil**, ampliando el ecosistema de IA sin restricciones éticas y permitiendo la generación y difusión de contenido potencialmente ilegal con mayor facilidad y anonimato (Barba Arteaga, 2024; Bigas Formatjé, 2023).

Estos ejemplos ilustran cómo la expansión de la IA generativa sin controles adecuados ha dado lugar a un entorno propicio para el abuso, la manipulación y la producción de material sensible, subrayando la urgencia de establecer marcos regulatorios y éticos robustos para estas tecnologías.

## 6. Desafíos y respuestas

El fenómeno de los chatbots de IA sin censura y su potencial influencia en la normalización de conductas sexuales violentas hacia menores plantea desafíos

complejos que requieren respuestas coordinadas desde múltiples ámbitos. Este apartado examina los vacíos regulatorios existentes, las lecciones aprendidas de las respuestas a la pornografía violenta, las iniciativas tecnológicas emergentes y los enfoques preventivos que se están desarrollando, con especial atención a la problemática específica de la simulación de abusos a menores.

### 6.1. Vacíos regulatorios actuales

El marco legal actual presenta importantes lagunas para abordar eficazmente los riesgos asociados a los chatbots sin censura y el contenido abusivo sintético:

- **Ausencia de tipificación específica:** En numerosas jurisdicciones, incluida España, existe una **carencia de leyes explícitas** que aborden el uso de tecnologías de IA para generar material de abuso sexual. **Aunque el 99% de las víctimas de deepfakes son mujeres, este fenómeno no se contempla aún como un delito específico en muchos códigos penales** (Bigas Formatjé, 2023; Barba Arteaga, 2024).
- **Paradoja legal en la simulación de abusos a menores:** Existe una **contradicción jurídica fundamental** en el caso de chatbots que simulan interacciones sexuales con menores. Mientras que la **legislación prohíbe claramente cualquier contenido visual de abuso infantil (incluso sintético)**, las **simulaciones textuales detalladas de abuso infantil mediante chatbots habitualmente caen en una zona gris legal** (Crawford & Smith, 2023). En muchas jurisdicciones, estas interacciones textuales no están tipificadas explícitamente como delito cuando no implican a menores reales ni generan imágenes, a pesar de su evidente potencial para normalizar comportamientos abusivos.
- **Heterogeneidad normativa:** A nivel internacional prevalece una **inconsistencia significativa** en cómo se abordan las simulaciones de abuso, con algunos países que las prohíben explícitamente bajo cualquier forma y otros que solo regulan el contenido visual o que involucra a menores reales (Martí Munné, 2023).

- **Vacío en la responsabilidad de los desarrolladores:** Las leyes actuales no establecen claramente las responsabilidades de las empresas y desarrolladores que crean y distribuyen chatbots capaces de simular interacciones abusivas con menores, especialmente cuando estos sistemas son deliberadamente diseñados "sin censura" (Thompson, 2023).
- **Dificultades probatorias:** La **identificación del perpetrador** constituye un obstáculo significativo debido al anonimato digital. Más complejo aún resulta establecer la intencionalidad de las interacciones con chatbots que simulan abusos: **¿constituyen fantasías protegidas por la libertad de expresión o son conductas preparatorias de delitos reales?** (La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea, 2024).
- **Extraterritorialidad del delito:** La naturaleza global de las plataformas de IA **complica la persecución de comportamientos inapropiados**, ya que los servidores, desarrolladores y usuarios pueden estar en diferentes jurisdicciones, creando refugios de impunidad para quienes explotan estos vacíos legales (Barba Arteaga, 2024; Lorao, 2024).

Estos vacíos regulatorios son particularmente preocupantes porque transmiten un **mensaje social ambiguo sobre la gravedad de simular abusos a menores, potencialmente contribuyendo a su normalización.**

## 6.2. Lecciones aprendidas de las respuestas a la pornografía violenta

Las estrategias desarrolladas para abordar la pornografía violenta proporcionan importantes lecciones aplicables al contexto específico de los chatbots que simulan abusos a menores:

- **Enfoque preventivo vs. reactivo:** La experiencia con la pornografía violenta demuestra que esperar a la materialización del daño para actuar resulta insuficiente. En el contexto de los chatbots, esto sugiere la necesidad de **implementar salvaguardas preventivas** antes de que se documenten casos de transferencia de conductas simuladas a abusos reales (Villanueva, 2024).

- **Progresión del consumo:** Los estudios sobre pornografía violenta han documentado patrones de **escalada en el consumo** donde **usuarios buscan progresivamente material más extremo**. Esta dinámica podría replicarse en interacciones con chatbots, donde la habituación podría llevar a simulaciones cada vez más extremas y potencialmente a la búsqueda de material visual que complemente estas fantasías (Leon et al., 2025; Lussier et al., 2021; Villanueva, 2024).
- **El papel de la fantasía en la conducta real:** La investigación sobre pornografía violenta ha establecido conexiones entre la exposición a fantasías violentas y cambios actitudinales que pueden manifestarse en comportamientos. Los chatbots que permiten simular abusos a menores podrían funcionar como **laboratorios virtuales** donde **se ensayan y refuerzan patrones conductuales problemáticos con un nivel de interactividad superior al de la pornografía tradicional** (Bhuptani, Cashin, et al., 2024; Wright et al., 2016).
- **Ineficacia del enfoque moral:** Se ha constatado que los enfoques basados exclusivamente en la condena moral del consumo de pornografía violenta resultan insuficientes. De manera similar, simplemente estigmatizar a quienes utilizan chatbots para simular abusos no aborda las raíces del problema ni ofrece alternativas efectivas para la prevención (Román-García et al., 2021).
- **Resistencia a la regulación:** La **industria pornográfica ha mostrado históricamente una gran capacidad para evadir regulaciones mediante cambios tecnológicos y jurisdiccionales**. El **mercado de chatbots sin censura** parece seguir patrones similares, con desarrolladores que **deliberadamente buscan circunvalar restricciones éticas, señalando la necesidad de estrategias más sofisticadas que la simple prohibición** (Thompson, 2023; Prichard et al., 2022).

### 6.3. Iniciativas tecnológicas y de plataformas

Frente a la creciente preocupación por la simulación de abusos a menores mediante chatbots, algunas plataformas e instituciones han comenzado a implementar medidas específicas:

- **Filtros contextuales avanzados:** A diferencia de los filtros léxicos tradicionales que bloquean palabras específicas, **algunas empresas están desarrollando sistemas de detección contextual capaces de identificar patrones conversacionales que sugieren simulaciones de abuso infantil, incluso cuando se utilizan eufemismos o lenguaje codificado** (Martí Munné, 2023).
- **Monitoreo proactivo:** Plataformas como Meta han implementado sistemas de **monitoreo proactivo** específicamente diseñados para detectar interacciones potencialmente abusivas con avatares que representan menores, tras incidentes documentados de este tipo de comportamiento (Horwitz, 2025; Meta Allows Facebook, Instagram AI Chatbots To Have Sex Talks With Children, 2025).
- **Prohibición explícita de la edad como parámetro en ciertos contextos:** Algunos desarrolladores han optado por eliminar completamente la posibilidad de crear avatares o personajes menores de edad en contextos conversacionales de naturaleza sexual o romántica (García, 2025).
- **Colaboración con entidades especializadas:** Iniciativas como NCMEC (National Center for Missing & Exploited Children) en Estados Unidos están estableciendo **protocolos de colaboración con empresas de IA para compartir patrones identificados de simulación de abusos y desarrollar contramedidas específicas** (La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea, 2024).
- **Detección de manipulación de sistemas:** Se están implementando sistemas para identificar intentos de eludir filtros mediante técnicas como "jailbreaking" (manipulación para eliminar restricciones éticas) o "prompt

engineering" (formulación estratégica de instrucciones para evadir protecciones) (Pacchioni et al., 2025).

Sin embargo, estos esfuerzos enfrentan limitaciones significativas:

- La **proliferación de modelos de código abierto que pueden ser modificados para eliminar restricciones éticas** (Merlio, 2025; Matatov et al., 2024)
- La tensión entre la **privacidad de las conversaciones** y la necesidad de monitoreo para detectar comportamientos problemáticos
- El desafío técnico de distinguir entre **narrativas ficticias legítimas** (como literatura o juegos de rol) y simulaciones abusivas (Martí Munné, 2023).
- La **carenza de estándares uniformes** sobre qué constituye simulación de abuso y cómo debe ser abordada

Un desafío particularmente relevante es la relación entre la libertad creativa y la prevención del abuso. A diferencia de la pornografía, donde las restricciones sobre contenido de abuso infantil están generalmente aceptadas, en el ámbito textual y conversacional existen debates más complejos sobre los límites de la expresión artística y la fantasía.

#### 6.4. La problemática específica de la simulación de abusos a menores

La simulación de abusos a menores mediante chatbots presenta características particulares que requieren consideración específica:

- **Función de entrenamiento conductual:** A diferencia del consumo pasivo de pornografía, la **interacción con chatbots permite un ensayo activo de técnicas de manipulación, seducción y coerción que podrían posteriormente transferirse a intentos de abuso real**. Investigaciones han documentado casos donde agresores sexuales utilizaron entornos virtuales para "practicar" antes de aproximarse a víctimas reales (Barba Arteaga, 2024; Inteligencia artificial y crimen organizado (AI Action Summit), 2024).

- **Normalización progresiva:** Las interacciones repetidas con chatbots que simulan ser menores pueden conducir a una desensibilización gradual donde el usuario percibe progresivamente la interacción sexual con menores como menos aberrante, potencialmente reduciendo inhibiciones internas contra tales comportamientos (Huesmann, 2007; Lussier et al., 2021).
- **Validación de fantasías preexistentes:** Para personas con atracciones sexuales preexistentes hacia menores, los chatbots sin censura pueden funcionar como sistemas de validación y refuerzo que legitiman estas inclinaciones como aceptables o inofensivas ("solo es un chatbot") (Barba Arteaga, 2024; Martí Munné, 2023).
- **Puerta de entrada a comunidades problemáticas:** La búsqueda de chatbots sin censura puede conectar a usuarios con comunidades online que normalizan el interés sexual por menores, facilitando la exposición a ideologías que racionalizan el abuso sexual infantil y potencialmente conduciendo a redes de intercambio de material ilegal (Matatov et al., 2024; Merlio, 2025).
- **Ausencia de consecuencias educativas:** A diferencia de las interacciones humanas, donde comportamientos inapropiados generalmente provocan rechazo o sanción social, los chatbots sin censura proporcionan un entorno donde estos comportamientos son aceptados o incluso reforzados, eliminando oportunidades de aprendizaje social correctivo (Johnson, 2025; Schibelsky, 2020).

Estas dinámicas únicas hacen que la simulación de abusos mediante chatbots pueda funcionar como catalizador de procesos de normalización más rápidos e intensos que los observados con otros medios, justificando la implementación de medidas preventivas específicas.

#### 6.5. Hacia un enfoque integral de prevención

Frente a la complejidad de este fenómeno emergente, se hace evidente la necesidad de un enfoque preventivo multidimensional:

- **Educación digital crítica:** Más allá de la alfabetización técnica, es necesario desarrollar **programas educativos que promuevan una comprensión ética de las interacciones con IA**, incluyendo la conciencia sobre cómo **ciertos patrones de uso pueden normalizar conductas dañinas** (Wolbers & Cubitt, 2025).
- **Intervención temprana:** Los programas de prevención deben incluir mecanismos de **identificación e intervención temprana** para personas que muestran patrones problemáticos de interacción con chatbots, similar a las iniciativas preventivas para potenciales agresores sexuales (Pacchioni et al., 2025).
- **Diseño ético por defecto:** La industria debe adoptar principios de "**ética por diseño**" donde las salvaguardas contra la simulación de abusos a menores sean elementos fundamentales de la arquitectura de los sistemas, no características añadidas posteriormente (Prichard et al., 2022; Martí Munné, 2023).
- **Colaboración entre sectores:** Es esencial la cooperación entre desarrolladores, educadores, profesionales de salud mental, autoridades y organizaciones de protección de menores para crear un **ecosistema preventivo** que aborde los múltiples factores que contribuyen a este fenómeno (La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea, 2024).
- **Investigación específica:** Existe una necesidad urgente de estudios rigurosos sobre los **efectos psicológicos, criminógenos y comportamentales** de las interacciones simuladas de abuso, para fundamentar políticas basadas en evidencia más que en reacciones emocionales o morales (Barba Arteaga, 2024; Wolbers & Cubitt, 2025).
- **Alternativas constructivas:** Es importante desarrollar aplicaciones de IA que canalicen el interés por la tecnología hacia **usos prosociales y creativos**, evitando que la búsqueda de experiencias transgresoras sea el principal atractivo de estas tecnologías (Pérez, 2023).

La particularidad de los chatbots como simuladores de abuso infantil requiere reconocer que nos encontramos ante un fenómeno sin precedentes históricos directos, donde la extrapolación de respuestas a otros tipos de contenido problemático puede resultar insuficiente. **La naturaleza interactiva, personalizada y potencialmente adictiva de estas tecnologías demanda respuestas igualmente innovadoras y adaptativas** (Martí Munné, 2023; Thompson, 2023).

En última instancia, abordar eficazmente este desafío requiere no solo medidas tecnológicas o legales, sino también un cuestionamiento más profundo sobre los valores sociales que permiten la objetivación sexual de menores en diversos contextos culturales, **reconociendo que los chatbots sin censura no crean estas tendencias, sino que amplifican y potencian dinámicas preexistentes en nuestra sociedad** (Umbach & Henry, 2025).

## 7. Conclusiones

### 7.1. Síntesis de hallazgos principales

La presente revisión sistemática ha examinado la influencia potencial de los chatbots de IA sin censura en la normalización de conductas sexuales violentas hacia menores, utilizando como marco analógico la evidencia empírica sobre pornografía violenta. **Los hallazgos principales revelan patrones preocupantes que merecen atención criminológica urgente.**

**Los chatbots de IA sin censura representan un nuevo paradigma en la potencial normalización de conductas abusivas**, distinguiéndose de fenómenos previos como la pornografía violenta por características fundamentales: **su interactividad bidireccional, la hiperpersonalización de experiencias, la co-creación de contenido abusivo, y su capacidad para funcionar como "entrenadores virtuales" de conductas problemáticas** (Barba Arteaga, 2024; Martí Munné, 2023; Thompson, 2023). Estas características crean un entorno único donde los mecanismos de normalización identificados en estudios sobre pornografía violenta podrían intensificarse significativamente.

**Los marcos teóricos criminológicos—particularmente la Teoría del Aprendizaje Social, la Teoría de la Desinhibición Online y los modelos de progresión delictiva—permiten comprender cómo estos sistemas pueden facilitar procesos de desensibilización, eliminar barreras morales contra el abuso, y potencialmente reducir inhibiciones que normalmente prevendrían conductas transgresoras** (Cid Moliné, 2023; Suler, 2004; Villanueva, 2024). Alarmantemente, los casos documentados de usuarios intentando crear chatbots que simulen menores en escenarios sexuales evidencian una demanda latente para explorar fantasías que serían inaceptables en cualquier otro contexto (Cox, 2024).

El análisis comparativo con la evidencia sobre pornografía violenta revela mecanismos compartidos de influencia: **la erotización de la violencia, la desensibilización progresiva, la validación de fantasías problemáticas, y la potencial transferencia de actitudes normalizadas hacia comportamientos reales** (Leon et al., 2025; Villanueva, 2024). Sin embargo, la **naturaleza interactiva y personalizada de los chatbots podría acelerar estos procesos y facilitar la escalada hacia conceptualizaciones cada vez más extremas del abuso** (Pacchioni et al., 2025).

**Particularmente preocupante resulta la confluencia de estos sistemas con el fenómeno del Abuso Sexual Basado en Imágenes (ASBI) en su versión sintética**, donde los chatbots pueden no solo normalizar conductas abusivas sino también facilitar la generación textual de contenido que detalle abusos a menores, operando en un vacío regulatorio y ético (Barba Arteaga, 2024; Maeng & Lee, 2022; Ramos-Zaga, 2024).

## 7.2. Implicaciones para la prevención criminológica

Las implicaciones de estos hallazgos para la prevención criminológica son profundas y multidimensionales, **requiriendo un enfoque integral que trascienda soluciones meramente técnicas o legales.**

Desde una perspectiva regulatoria, **es imperativo desarrollar marcos normativos específicos que aborden las simulaciones de abuso hacia menores mediante chatbots**, superando la paradoja legal actual donde muchas jurisdicciones penalizan el contenido visual abusivo sintético pero no las

simulaciones textuales equivalentes (Barba Arteaga, 2024; Bigas Formatjé, 2023). La tipificación de estas conductas como delitos sexuales, no meras infracciones de privacidad, resultaría más coherente con su potencial criminógeno.

**El principio de "ética por diseño" debe adoptarse como estándar industrial**, incorporando salvaguardas multicapa que detecten y bloqueen activamente patrones conversacionales consistentes con abusos a menores, incluso cuando se utilice lenguaje eufemístico o codificado (Martí Munné, 2023; Pacchioni et al., 2025). Las empresas desarrolladoras deben asumir responsabilidad legal y ética por las capacidades de sus sistemas, implementando auditorías regulares y transparentes.

**Las intervenciones preventivas deben operar simultáneamente en múltiples niveles.** A nivel primario, programas educativos que integren alfabetización digital crítica con educación afectivo-sexual basada en el consentimiento y el respeto pueden reducir la demanda de este tipo de contenido. A nivel secundario, sistemas de identificación temprana e intervención con usuarios que muestran patrones de interacción problemáticos podrían prevenir la escalada. A nivel terciario, programas especializados para usuarios que ya han desarrollado patrones abusivos en entornos virtuales podrían reducir el riesgo de transferencia a comportamientos reales (Barba Arteaga, 2024; Wolbers & Cubitt, 2025).

Particularmente prometedor resulta **el desarrollo de intervenciones digitales preventivas como chatbots específicamente diseñados para interrumpir trayectorias de escalada**, aumentando la percepción de riesgo y facilitando el acceso a recursos de apoyo para individuos preocupados por sus impulsos o comportamientos (Martí Munné, 2023). Estas herramientas podrían aplicar principios de la terapia cognitivo-conductual para abordar distorsiones cognitivas que facilitan la desconexión moral.

### 7.3. Limitaciones del estudio y analogía utilizada

Esta revisión presenta limitaciones inherentes que deben considerarse al interpretar sus conclusiones. **La más significativa radica en la escasez de investigación empírica directa sobre el impacto específico de los chatbots sin censura en la normalización de conductas abusivas**, obligando a establecer

extrapolaciones desde campos relacionados como la pornografía violenta. Aunque la analogía proporciona un marco conceptual valioso, los mecanismos de influencia podrían diferir en aspectos no contemplados debido a la naturaleza única de las interacciones con IA.

Metodológicamente, **la heterogeneidad de los estudios analizados, muchos de ellos correlacionales, limita la capacidad para establecer relaciones causales definitivas**. La rapidez con que evoluciona la tecnología de IA también presenta un desafío para cualquier revisión sistemática, potencialmente dejando fuera desarrollos recientes que podrían alterar las dinámicas analizadas.

**La analogía con la pornografía violenta, aunque fundamentada en similitudes conceptuales sólidas, no captura completamente aspectos únicos de los chatbots** como su capacidad para simular relaciones interpersonales complejas o para adaptar dinámicamente sus respuestas a las preferencias del usuario (Bhuptani, Cashin, et al., 2024; Leon et al., 2025). Estas diferencias podrían implicar mecanismos de influencia no contemplados en la literatura sobre pornografía.

#### 7.4. Futuras líneas de investigación

Este trabajo abre múltiples vías para investigación futura, esenciales para desarrollar respuestas basadas en evidencia ante este fenómeno emergente.

**Resulta prioritario conducir estudios longitudinales que examinen específicamente la relación entre el uso prolongado de chatbots sin restricciones éticas y cambios actitudinales o comportamentales hacia la violencia sexual.** Estudios experimentales controlados podrían evaluar si la exposición a estos sistemas modifica percepciones sobre el consentimiento o la gravedad del abuso infantil (Pacchioni et al., 2025; Wolbers & Cubitt, 2025).

Metodológicamente innovador sería **desarrollar y validar métricas cuantitativas para evaluar el riesgo criminógeno de diferentes modelos de IA**, midiendo factores como su tendencia a la auto-inhibición ante prompts ilegales o su capacidad para generar narrativas que justifiquen la violencia sexual (Prichard et al., 2022).

**La intersección entre chatbots sin censura y otros fenómenos digitales problemáticos, como el ciber-grooming o la creación de deepfakes, merece atención específica.** Particularmente relevante resulta investigar si existe una progresión desde interacciones textuales simuladas hacia la búsqueda o creación de contenido visual abusivo (Barba Arteaga, 2024; Lorao, 2024).

En el ámbito aplicado, **el desarrollo y evaluación rigurosa de intervenciones preventivas basadas en IA**, como chatbots terapéuticos o sistemas de detección temprana de patrones problemáticos, podría traducir los hallazgos teóricos en herramientas prácticas de prevención criminológica (La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea, 2024).

Finalmente, **estudios sobre la efectividad comparativa de diferentes enfoques regulatorios internacionales permitirían identificar mejores prácticas para abordar este desafío transnacional**, reconociendo que ninguna jurisdicción puede enfrentarlo aisladamente (Inteligencia artificial y crimen organizado (AI Action Summit, 2024)).

La complejidad ética, técnica y social de este fenómeno demanda un esfuerzo investigador multidisciplinar que integre perspectivas criminológicas, psicológicas, tecnológicas y legales para comprender plenamente sus implicaciones y desarrollar estrategias preventivas efectivas que protejan a los menores sin obstaculizar el desarrollo responsable de la inteligencia artificial.

## 8. Referencias (APA 7 edición)

*AI chatbots and companions – risks to children and young people | eSafety Commissioner.* (2025, febrero 18). <https://www.esafety.gov.au/newsroom/blogs/ai-chatbots-and-companions-risks-to-children-and-young-people>

Amantegui Guezala, A. (2022, enero 21). *Crean una Inteligencia Artificial capaz de mantener conversaciones naturales.. Y los usuarios le hacen bullying.* La Vanguardia. <https://www.lavanguardia.com/tecnologia/20220121/7999707/crean-inteligencia-artificial-capaz-mantener-conversaciones-naturales-pmv.html>

Barba Arteaga, C. (2024). Deepfakes sexuales: Impacto, prevención y perspectivas de género en el entorno digital. *Miguel Hernández Communication Journal, 15*, 229-244.

Bardhan, A. (2022, enero 18). *Men Are Creating AI Girlfriends and Then Verbally Abusing Them.* Futurism. <https://futurism.com/chatbot-abuse>

Becerra-Alfonso, Y. A. (2015). Trastornos del dolor sexual femenino: Una revisión de su definición, etiología y prevalencia. *Medicas UIS, 28(3)*, 267-272.

Bhuptani, P. H., Cashin, D., & Orchowski, L. M. (2024). The Impact of Violent Pornography on Sexual Coercive Behaviors Among College Men: A Prospective Examination. *Journal of Interpersonal Violence, 38*(1), 8862605241299442. <https://doi.org/10.1177/08862605241299442>

Bhuptani, P. H., Kenney, S. R., Napper, L. E., & Orchowski, L. M. (2024). Pornography Use, Perceived Peer Norms, and Attitudes Toward Women: A Study of College Men. *American journal of sexuality education, 19(3)*, 280-301. <https://doi.org/10.1080/15546128.2023.2233414>

Bigas Formatjé, N. (2023). *El 99% del deepfake dirigido contra las mujeres.* Universitat Oberta de Catalunya. <https://www.uoc.edu/es/news/2023/265-deepfakes-pornograficos-cuando-IA-desnuda-tu-intimidad-vulnera-tus-derechos>

Cid Moliné, J. (with Larrauri, E.). (2023). *Teorías criminológicas: Explicación y prevención de la delincuencia* (2ª Edición). La Ley Soluciones Legales, S.A.

Colton, E., & News, F. (2023, julio 22). *AI could «normalize» child sex abuse as graphic images erupt online: Experts.* <https://www.foxnews.com/tech/artificial-intelligence-could-normalize-child-sexual-abuse-graphic-images-erupt-online-experts>

Cox , J. (2024, octubre 8). *Hacked ‘AI Girlfriend’ Data Shows Prompts Describing Child Sexual Abuse.* 404 Media. <https://www.404media.co/hacked-ai-girlfriend-data-shows-prompts-describing-child-sexual-abuse-2/>

Crawford, A., & Smith, T. (2023, junio 29). *IA: El comercio ilegal de imágenes de abuso sexual infantil generadas por inteligencia artificial.* BBC News Mundo. <https://www.bbc.com/mundo/articles/c1vzyevl0nro>

Fagan, J. A. (1986). [Review of *Child Sexual Abuse: New Theory and Research*, por D. Finkelhor]. *The Journal of Criminal Law and Criminology* (1973-), 77(2), 477-481. <https://doi.org/10.2307/1143341>

García, C. (2025). *Hombres solitarios están creando novias con IA y desatan su ira violenta contra ellas, según expertos.* El Imparcial (Edición México). <https://www.elimparcial.com/locurioso/2025/02/17/hombres-solitarios-estan-creando-novias-con-ia-y-desatan-su-ira-violenta-contra-ellas-segun-expertos/>

Generación expuesta: Jóvenes frente a la violencia sexual digital. (2024). *Centro Reina Sofía de Fad Juventud.* Recuperado 8 de mayo de 2025, de <http://www.centroreinasofia.org/publicacion/generacion-expuesta-jovenes-violencia-sexual-digital/>

Henry, N., Witt ,Alice, & and Vasil, S. (2024). A ‘design justice’ approach to developing digital tools for addressing gender-based violence: Exploring the possibilities and limits of feminist chatbots. *Information, Communication & Society*, 0(0), 1-24. <https://doi.org/10.1080/1369118X.2024.2363900>

Hernández Flórez, N. E., Acevedo Villalobos, E. P., & Espeleta Maradei, M. (2013). Revisión teórica en torno a la discusión sobre trastornos sexuales. *Psicoespacios: Revista virtual de la Institución Universitaria de Envigado*, 7(10), 291-320.

Horwitz, J. (2025). *Meta's 'Digital Companions' Will Talk Sex With Users—Even Children.* The Wall Street Journal. <https://www.wsj.com/tech/ai/meta-ai-chatbots-sex-a25311bf>

Huesmann, L. R. (2007). The Impact of Electronic Media Violence: Scientific Theory and Research. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, 41(6 Suppl 1), S6-13. <https://doi.org/10.1016/j.jadohealth.2007.09.005>

*Inteligencia artificial y crimen organizado (AI Action Summit).* (2024). El Paccto 2.0 EU-LAC Partnership on justice and security. <https://www.fiap.gob.es/wp-content/uploads/2024/11/ELPACCTO2-IAyCrimen-ES.pdf>

Isom, D., Blackwell, B., & Hoyle, J. (2015, noviembre 19). *The Empirical Status of General Strain Theory: A Meta-Analysis: Presentation at the Annual Meeting of the American Society of Criminology.*

*IWF 2024 Annual Data & Insights Report.* (2024). Internet Watch Foundation. <https://www.iwf.org.uk/annual-data-insights-report-2024/>

Johnson, D. (2025, marzo 5). Anorexia coaches, self-harm buddies and sexualized minors: How online communities are using AI chatbots for harmful behavior. *CyberScoop.* <https://cyberscoop.com/graphika-ai-chatbots-harmful-behavior-character-ai/>

*La inteligencia artificial y la lucha contra la explotación y el abuso sexual de menores en línea.* (2024). Know2Protect. Homeland Security Investigations. <https://www.dhs.gov/sites/default/files/202>

*Las mejores 145 No apto para menores herramientas de AI.* (2025). Toolify.ai. <https://www.toolify.ai/es/category/nsfw>

Leon, C. M., Quiñonez-Toral, T., & Aizpurua, E. (2025). From Pornography Consumption to Sexually Violent Practices: Uncovering the Hidden Influence of Sexual Norms. *Behavioral Sciences*, 15(3), Article 3. <https://doi.org/10.3390/bs15030243>

Lorao, D. (2024, octubre 18). *Preocupación por el aumento de imágenes de abuso sexual infantil generadas por IA.* Artículo 14. <https://www.articulo14.es/violencia-contra-las-mujeres/preocupacion-por-el-aumento-de-imagenes-de-abuso-sexual-infantil-generadas-por-ia-20241018.html>

*Los 8 mejores chatbots de IA sin censura en 2025.* (2025, febrero 22). [AIMOJO]. <https://aimojo.io/es/uncensored-ai-chatbots/>

Loutzenhiser, L., Arrighi, F., & Rosenfeld, B. (2024). The association between pornography use and sexual offending in individuals with a history of sex offenses: A meta-analysis. *Aggression and Violent Behavior*, 78, 101980. <https://doi.org/10.1016/j.avb.2024.101980>

Lussier, P., McCuish, E., & James, J. (2021). *Sexual Deviance*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119771401.ch2>

Maeng, W., & Lee, J. (2022). Designing and Evaluating a Chatbot for Survivors of Image-Based Sexual Abuse. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1-21. <https://doi.org/10.1145/3491102.3517629>

Martí Munné, C. (2023, agosto 14). Inteligencia Artificial (IA) y Material de Explotación Sexual Infantil (MESI). *AngelBlau*. <https://angelblau.com/inteligencia-artificial-y-mesi/>

Matatov, H., Quéré, M. A. L., Amir, O., & Naaman, M. (2024). *Examining the Prevalence and Dynamics of AI-Generated Media in Art Subreddits* (arXiv:2410.07302). arXiv. <https://doi.org/10.48550/arXiv.2410.07302>

Merlio. (2025). *How to Bypass AI Content Filters Responsibly and Ethically*. Merlio. <https://merlio.app>

*Meta Allows Facebook, Instagram AI Chatbots To Have Sex Talks With Children: Report.* (2025). NDTV. <https://www.ndtv.com/world-news/meta-allows-facebook-instagram-ai-chatbots-to-have-sex-talks-with-children-report-8274354>

News, L. (2025, abril 28). La IA de Meta mantiene chats sexuales con menores, según WSJ. *LISA News*. <https://www.lisanews.org/actualidad/la-ia-ai-de-meta-mantiene-chats-sexuales-con-menores-segun-wsj/>

*NSFW.* (2025, mayo 7). Cambriadge Dictionary.  
<https://dictionary.cambridge.org/dictionary/english/nsfw>

Pacchioni, F., Flutti, E., Caruso, P., Fregna, L., Attanasio, F., Passani, C., Colombo, C., & Travaini, G. (2025). Generative AI and criminology: A threat or a promise? Exploring the potential and pitfalls in the identification of Techniques of Neutralization (ToN). *PLOS One*, 20(4), e0319793.  
<https://doi.org/10.1371/journal.pone.0319793>

Pastor, J. (2013, julio 11). *Negobot, un desarrollo español a la caza de pedófilos en los chats.* Xataka. <https://www.xataka.com/otros/negobot-un-desarrollo-espanol-a-la-caza-de-pedofilos-en-los-chats>

Peckham, J. (2025, abril 28). *Meta AI Capable of Having Sexual Conversations With Minors.* PC MAG. <https://www.pc当地.com/news/meta-ai-capable-of-having-sexual-conversations-with-minors>

Pérez, H. (2023, abril 16). *Sexting con robots: La IA también puede hablar sucio y a muchos les encanta.* DiarioBitcoin. <https://www.diariobitcoin.com/ia/sexting-con-robots-ia-tambien-puede-hablar-sucio-y-a-muchos-les-encanta/>

Prichard, J., Scanlan, J., Watters, P., & Wortley, R. (2022). *Online messages to reduce users' engagement with child sexual abuse material: A review of relevant literature for the rethink Chatbot.* ResearchGate.  
[https://www.researchgate.net/publication/361464361\\_Online\\_messages\\_to\\_reduce\\_users'\\_engagement\\_with\\_child\\_sexual\\_abuse\\_material\\_a\\_review\\_of\\_relevant\\_literature\\_for\\_the\\_rethink\\_Chatbot](https://www.researchgate.net/publication/361464361_Online_messages_to_reduce_users'_engagement_with_child_sexual_abuse_material_a_review_of_relevant_literature_for_the_rethink_Chatbot)

Ramos-Zaga, F. (2024). Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial. *Derecho Global. Estudios sobre Derecho y Justicia*, 9(27), Article 27. <https://doi.org/10.32870/dgedj.v9i27.754>

Rodrigues, C. (2025). *Reddit denuncia experimento no autorizado con IA: bots fingieron ser víctimas para manipular a usuarios.* RPP.  
<https://rpp.pe/tecnologia/mas-tecnologia/reddit-denuncia-experimento-no-autorizado-con-ia-bots-fingieron-ser-victimas-para-manipular-a-usuarios-noticia-1631448>

Román-García, Ó., Bacigalupe, A., Vaamonde-García, C., Román-García, Ó., Bacigalupe, A., & Vaamonde-García, C. (2021). Relación de la pornografía mainstream con la salud sexual y reproductiva de los/las adolescentes. Una revisión de alcance. *Revista Española de Salud Pública*, 95. [https://scielo.isciii.es/scielo.php?script=sci\\_abstract&pid=S1135-57272021000100182&lng=es&nrm=iso&tlang=es](https://scielo.isciii.es/scielo.php?script=sci_abstract&pid=S1135-57272021000100182&lng=es&nrm=iso&tlang=es)

Saz Rubio, M. (2024, diciembre 4). *El maltrato verbal hacia las mujeres en las redes se traslada a la inteligencia artificial.* La Vanguardia. <https://www.lavanguardia.com/vivo/relaciones/20241204/10164435/maltrato-verbal-mujeres-redes-traslada-inteligencia-artificial-pmv.html>

Schibelsky, L. (2020). Children's Online Safety: Prospecting Chatbots for Tackling Online Abuse. *Knowledge Media Institute. Centre for Policing Research and Learning.*

Singh, S., & Nambiar, V. (2024). Role of Artificial Intelligence in the Prevention of Online Child Sexual Abuse: A Systematic Review of Literature. *Journal of Applied Security Research*, 19(4), 586-627. <https://doi.org/10.1080/19361610.2024.2331885>

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 7(3), 321-326. <https://doi.org/10.1089/1094931041291295>

swissinfo.ch, S. W. I. (2025, febrero 2). El R.Unido quiere ilegalizar el uso de IA para crear material con fines de abuso sexual. *SWI swissinfo.ch.* <https://www.swissinfo.ch/spa/el-r.unido-quiere-ilegalizar-el-uso-de-ia-para-crear-material-con-fines-de-abuso-sexual/88814804>

Thompson, S. A. (2023a, julio 2). Uncensored Chatbots Provoke a Fracas Over Free Speech. *The New York Times.* <https://www.nytimes.com/2023/07/02/technology/ai-chatbots-misinformation-free-speech.html>

Thompson, S. A. (2023b, julio 24). *Los chatbots sin censura provocan caos en torno a la libertad de expresión.* infobae. <https://www.infobae.com/america/the->

[new-york-times/2023/07/24/los-chatbots-sin-censura-provocan-caos-en-torno-a-la-libertad-de-expresion/](https://new-york-times/2023/07/24/los-chatbots-sin-censura-provocan-caos-en-torno-a-la-libertad-de-expresion/)

Umbach, R., & Henry, N. (2025, abril 30). *Prevalence and Impacts of Image-Based Sexual Abuse Victimization: A Multinational Study*. ResearchGate. <https://doi.org/10.48550/arXiv.2503.04988>

Villanueva, C. M. (2024). Consumo de pornografía y normalización de conductas violentas en las relaciones sexuales de los jóvenes. *Atlánticas. Revista Internacional de Estudios Feministas*, 9(1), Article 1. <https://doi.org/10.17979/arief.2024.9.1.9401>

Wolbers, H., & Cubitt, T. (2025, enero 24). *Artificial intelligence and child sexual abuse: A rapid evidence assessment*. ResearchGate. <https://doi.org/10.52922/ti77802>

Wright, P. J., Tokunaga, R. S., & Kraus, A. (2016). A Meta-Analysis of Pornography Consumption and Actual Acts of Sexual Aggression in General Population Studies. *Journal of Communication*, 66(1), 183-205. <https://doi.org/10.1111/jcom.12201>

Yang, D., & Youn, G. (2012). *Effects of Exposure to Pornography on Male Aggressive Behavioral Tendencies*. 5(1). <https://doi.org/10.2174/1874350101205010001>

