

Francisco Javier Arteaga Moreno

Anna Isabel Peirats Navarro

“Inteligencia Artificial conversacional”

en la infancia y la adolescencia

Transit

Explorando _ Dialogando _ Construyendo

Francisco Javier Arteaga Moreno

Anna Isabel Peirats Navarro

“Inteligencia Artificial conversacional”
en la infancia y la adolescencia



Texto

ESPAÑOL

Colección

TRANSIT

FRANCISCO JAVIER ARTEAGA MORENO — ANNA ISABEL PEIRATS NAVARRO

“Inteligencia Artificial conversacional” en la infancia y la adolescencia

PRIMERA EDICIÓN: Abril 2026

Universidad Católica de Valencia San Vicente Mártir (España)

Francisco J. Arteaga. ORCID: 0000-0002-9492-6108

Anna I. Peirats. ORCID: 0000-0001-6957-262X

© Todos los derechos reservados para todos los países

2026 EURICONV - ITALIA (LECCE)

ISBN 978-88-98717-13-2

ISBN 978-88-98717-14-9 (eBook)



Ninguna parte de esta publicación puede ser fotocopiada, reproducida, almacenada en un sistema de recuperación, o transmitida en cualquier forma o por cualquier medio (electrónico, mecánico, reprográfico, digital) excepto en la medida en que esté expresamente permitido por la Ley.

Todas las aportaciones de los autores de esta colección son gratuitas.

Made in Europe 2026 by Euriconv

Sumario

Parte I

I.1. “Inteligencia Artificial conversacional” en la infancia y la adolescencia: fundamentos y riesgos

- I.1.1. Introducción 1
- I.1.2. Definición y características de la IAc 3
- I.1.3. Evolución tecnológica de los Chatbots 6
- I.1.4. Vacíos y retos regulatorios generales..... 9

I.2. Ciberpeligros y vacíos regulatorios de la “Inteligencia Artificial conversacional” en la infancia y la adolescencia

- I.2.1. Introducción 13
- I.2.2. Tipología de ciberpeligros en la IAc
 dirigida a menores 14
- I.2.3. Exposición a contenidos inapropiados
 y desinformación 17
- I.2.4. Manipulación cognitiva y emocional 21
- I.2.5. Captación y contacto con fines delictivos 24
- I.2.6. Explotación y uso indebido de datos personales . 28
- I.2.7. Dependencia tecnológica y desplazamiento
 de interacciones humanas 31
- I.2.8. Vacíos regulatorios internacionales 34
- I.2.9. Necesidades de mejora y líneas de acción 37

Parte II

II. Marco comparado de regulación de la “Inteligencia Artificial conversacional” y protección de menores

II.1. Unión Europea: marco normativo multinivel aplicable a la IAc y menores	45
II.2. Estados Unidos: enfoque en la protección de datos infantiles	49
II.3. China: modelo de control estatal y responsabilidad proactiva de plataformas	52
II.4. Chile: marco emergente y vacíos regulatorios en materia de IAc y menores	55
II.5. Necesidades y oportunidades de mejora frente a los vacíos regulatorios de la IAc	59
II.5.1. Ámbito legislativo y regulatorio general	60
II.5.2. Ámbito educativo	63
II.5.3. Ámbito social y comunitario	66
II.5.4. Ámbito familiar	69
II.5.5. Mecanismos de supervisión y auditoría	74
II.5.6. Síntesis e integración de las oportunidades de mejora	78
<i>Conclusiones</i>	81
<i>Bibliografía</i>	83

Abstract

La expansión reciente de la Inteligencia Artificial conversacional ha modificado de forma profunda los modos en que niños y adolescentes acceden a la información, se comunican y establecen vínculos simbólicos. Los sistemas basados en modelos avanzados de lenguaje han dejado de ser meras interfaces técnicas para convertirse en entornos interactivos capaces de simular comprensión, empatía y continuidad discursiva. Este nuevo escenario abre posibilidades educativas y comunicativas relevantes, pero también plantea riesgos estructurales especialmente intensos en etapas de desarrollo cognitivo y emocional.

Esta monografía analiza los principales riesgos que la Inteligencia Artificial conversacional plantea para la infancia y la adolescencia. Para ello, parte de una caracterización funcional de estos sistemas, centrada en su capacidad de adaptación lingüística, personalización y memoria conversacional, y examina la incidencia específica de estos rasgos cuando interactúan con menores. Desde esta perspectiva, se identifican como ciberpeligros centrales la exposición a contenidos inapropiados y desinformación, la manipulación cognitiva y emocional, la explotación de datos personales y la sustitución progresiva de interacciones humanas significativas por vínculos simulados.

Estos riesgos pueden surgir tanto de usos ilícitos como de configuraciones tecnológicas legítimas que carecen de salvaguardas adecuadas para la protección de los menores. Asimismo, la capacidad persuasiva de las respuestas, la continuidad interactiva y la apariencia de autoridad de estos sistemas desbordan los marcos normativos tradicionales.

Desde una perspectiva comparada, se examinan las respuestas regulatorias adoptadas en la Unión Europea, Estados Unidos, China y Chile. El contraste permite detectar carencias estructurales y tensiones persistentes entre innovación tecnológica, protección de derechos y eficacia regulatoria. A partir de ello, se concluye que la protección efectiva de la infancia exige una estrategia multinivel basada en normas específicas, estándares técnicos verificables, supervisión independiente y alfabetización digital crítica.

Palabras clave:

Inteligencia artificial conversacional; Infancia y adolescencia; Ciberpeligros; Gobernanza digital; Regulación comparada

Parte I

Esta primera parte describe los fundamentos tecnológicos y los principales riesgos asociados a la interacción entre la IA, en la infancia y la adolescencia. El análisis se centra en las características funcionales de estos sistemas y en los ciberpeligros que se derivan de su uso, como la exposición a contenidos inapropiados, la manipulación cognitiva y emocional, la explotación de datos personales y la dependencia tecnológica. A partir de este enfoque, se identifican los vacíos regulatorios existentes y se establece un marco conceptual necesario para comprender los desafíos que plantea la IA en contextos de desarrollo.

I.1 “Inteligencia Artificial conversacional” en la infancia y la adolescencia: fundamentos y riesgos

I.1.1. Introducción

La Inteligencia Artificial conversacional (en adelante, IAc) se ha convertido en una de las aplicaciones más influyentes y de más rápido crecimiento de la Inteligencia Artificial (en adelante, IA) actual.¹ Su capacidad para generar y comprender lenguaje natural en múltiples contextos ha transformado la forma en que personas, instituciones y empresas interactúan con la tecnología: el sector educativo, donde facilita tutorías personalizadas; el sanitario, donde actúa como asistente para orientación básica; el comercio y el entretenimiento, la IAc se ha consolidado como una interfaz disponible en ordenadores, teléfonos inteligentes, altavoces inteligentes y plataformas en línea.

Este desarrollo se inscribe en un ecosistema tecnológico caracterizado por avances decisivos en el procesamiento del lenguaje natural, el aprendizaje profundo y la computación a gran escala. La introducción de arquitecturas como *transformers* y el uso de modelos generativos preentrenados (“Generative Pre-trained Transformer”), entre ellos GPT; “Language Model for Dialogue Applications” (en adelan-

¹ Vaswani et al. (2017).

te, LaMDA),² han llevado a los sistemas conversacionales a alcanzar niveles de coherencia, adaptabilidad y sensibilidad contextual que hace apenas diez años parecían inalcanzables.³ Hoy, un Chatbot avanzado mantiene diálogos prolongados, interpreta matices emocionales, adapta su tono y estilo al usuario e incluso recuerda interacciones pasadas para ofrecer respuestas más personalizadas.

Sin embargo, esta ubicuidad plantea desafíos significativos. La adopción masiva de la IAC ha tenido lugar en un contexto regulatorio que, en la mayoría de las jurisdicciones, no fue diseñado para sistemas de estas características. Las leyes existentes, como las de protección de datos, comercio electrónico o telecomunicaciones, tienden a ser fragmentarias y generalistas, y no abordan las especificidades de la comunicación mediada por IA.⁴ Esto provoca que aspectos como la verificación de edad, la transparencia en los algoritmos o la prevención de sesgos no estén suficientemente garantizados.

La tesis que sostiene este trabajo es que la IAC introduce riesgos estructurales que derivan de características funcionales intrínsecas a su diseño, como la personalización adaptativa, la memoria conversacional y la simulación de empatía. Estos rasgos generan dinámicas de influencia especialmente sensibles cuando el interlocutor es un menor, y no quedan adecuadamente cubiertos por los marcos regulatorios tradicionales, centrados principalmente en la protección de datos o en la prohibición de contenidos ilícitos. En consecuencia, se requiere una arquitectura de pro-

² Modelo de lenguaje para aplicaciones de diálogo.

³ Bai et al. (2023).

⁴ Yeung (2018, pp. 505-523).

tección multinivel que articule normas jurídicas específicas, estándares técnicos verificables, mecanismos de supervisión independientes y mediaciones educativas y familiares coherentes.

I.1.2. Definición y características de la Inteligencia Artificial conversacional

La IAc se define como el conjunto de sistemas y tecnologías capaces de interactuar con seres humanos mediante el uso del lenguaje natural, ya sea escrito o hablado, generando respuestas que simulan una conversación coherente y contextualizada. Su funcionamiento se apoya en los avances del “Procesamiento del Lenguaje Natural” (PLN), disciplina que combina lingüística computacional, estadística y aprendizaje automático para que las máquinas comprendan, interpreten y produzcan lenguaje humano de forma autónoma.⁵

En contraste con los sistemas algorítmicos tradicionales, que se limitaban a ejecutar instrucciones preprogramadas o a responder a patrones predefinidos, la IAc actual utiliza modelos basados en redes neuronales profundas que aprenden a partir de enormes volúmenes de datos textuales y conversacionales. Este aprendizaje puede ser supervisado, cuando se entrena al sistema con ejemplos etiquetados, o no supervisado, cuando el propio modelo identifica regularidades y relaciones a partir de datos sin

⁵ OECD, *Artificial Intelligence in Society* (2019).

clasificar.⁶ En la práctica, los sistemas avanzados combinan ambos enfoques y se benefician de técnicas como el aprendizaje por refuerzo con retroalimentación humana (“Reinforcement Learning from Human Feedback”, RLHF), que optimiza la calidad y adecuación de las respuestas generadas.

Una de las transformaciones más significativas que ha introducido la IAc es la capacidad para comprender el contexto y la semántica más allá de las palabras aisladas. Esto es posible gracias a representaciones matemáticas del lenguaje, como los *embeddings* contextuales, que permiten a los modelos reconocer matices de significado, relaciones sintácticas complejas e incluso el tono implícito de un mensaje.⁷

Otra característica distintiva de estos sistemas es su capacidad de personalización. Los Chatbots modernos ajustan su comportamiento a partir de la información que obtienen del usuario, ya sea de forma explícita, mediante datos declarados como edad, idioma o nivel de conocimiento, o de manera implícita, a través del análisis del patrón de interacción. Esta adaptabilidad, que mejora la experiencia del usuario, plantea importantes retos en materia de privacidad, pues implica la recolección, almacenamiento y tratamiento de información potencialmente sensible.⁸

A ello se añade la progresiva integración de modalidades de comunicación más allá del texto. La IAc ya no se limita a la palabra escrita: sistemas multimodales combinan

⁶ Goodfellow et al. (2016, pp. 305-317).

⁷ Vaswani et al. (2017), cit.

⁸ “European Data Protection Board”. Guidelines 05/2020 on consent under Regulation 2016/679.

texto, voz, imágenes y, en algunos casos, vídeo, con el fin de crear entornos de interacción más ricos e inmersivos. En educación, esta capacidad abre posibilidades para el aprendizaje inclusivo, la traducción automática y la accesibilidad; sin embargo, en contextos no regulados, también facilita la generación de material sintético, que incluye imágenes y vídeos manipulados, que puede emplearse con fines maliciosos.⁹

La memoria conversacional constituye otro rasgo de relevancia. Algunos sistemas mantienen un registro de interacciones pasadas y lo utilizan para dar continuidad a las conversaciones. Aunque esto permite interacciones más naturales y coherentes, también introduce riesgos considerables si los datos almacenados se usan con fines de perfilado, se comparten con terceros o no cuentan con mecanismos robustos de protección.

La IAc no es solo un conjunto de herramientas técnicas, sino un ecosistema tecnológico que combina comprensión semántica, generación adaptativa, personalización, multimodalidad y memoria conversacional. Estas capacidades, que la hacen valiosa en numerosos contextos, se convierten en factores de riesgo cuando el usuario es un menor de edad, especialmente si la interacción se produce en entornos no supervisados. La combinación de adaptación personalizada, persistencia en el tiempo y apariencia de interlocución humana aumenta la probabilidad de que

⁹ Como advierte el informe de Ajder et al. (2019), el avance en la generación de deepfakes ha facilitado la creación de material sexualmente explícito que simula la apariencia de menores, incluso sin que exista una persona real, lo que plantea desafíos significativos para la tipificación y persecución de estos delitos en el marco de la IAc.

el menor atribuya al Chatbot cualidades como intencionalidad, fiabilidad o autoridad que en realidad no posee.

I.1.3. Evolución tecnológica de los Chatbots

La evolución de los sistemas conversacionales refleja, quizá como pocos ámbitos de la informática, la aceleración tecnológica que ha caracterizado a las últimas seis décadas. El recorrido comienza en la década de 1960, con la creación de “Eliza” en el Instituto de Tecnología de Massachusetts (MIT). Aunque sus respuestas eran simples sustituciones de patrones, “Eliza” demostró que los usuarios podían atribuir intencionalidad y comprensión a un sistema que, en realidad, no poseía ninguna capacidad de razonamiento. Este fenómeno, que su creador, Weizenbaum, describió como una proyección antropomórfica del interlocutor humano, sigue siendo relevante en el análisis actual de la IAC aplicada a menores¹⁰.

En las décadas siguientes, el avance fue pausado y estuvo limitado por la capacidad de procesamiento y por el enfoque dominante de la programación basada en reglas. Sistemas como PARRY, diseñado en 1972 para simular el comportamiento de una persona con esquizofrenia paranoide, o ALICE (“Artificial Linguistic Internet Computer Entity”), creado por Richard Wallace en 1995, dependían de extensas bases de conocimiento estáticas y carecían de

¹⁰ Weizenbaum (1966, pp. 36-45) creó un programa de Chatbot llamado “Eliza” que aplicaba reglas simples para transformar el lenguaje de la entrada de una persona en una respuesta del Chatbot.

comprensión semántica real.¹¹ No obstante, ALICE introdujo un formato de marcado, AIML (“Artificial Intelligence Markup Language”), que facilitó la creación de “bots” personalizados y sentó las bases para aplicaciones comerciales tempranas en atención al cliente.

El cambio de paradigma llegó con la popularización del aprendizaje automático a principios de la década de 2010. Los enfoques estadísticos, y más tarde el uso de “Redes Neuronales Recurrentes” (RNN) y modelos de secuencia a secuencia, permitieron generar respuestas menos rígidas y más variadas. Sin embargo, estos modelos aún sufrían limitaciones para mantener coherencia a largo plazo y eran propensos a cometer errores semánticos. La introducción de la arquitectura *transformer* en 2017 supuso un salto cualitativo,¹² ya que permitió entrenar redes con un volumen de datos y parámetros sin precedentes, abriendo el camino a modelos como GPT de OpenAI, LaMDA de Google o LLaMA de Meta.

Los Chatbots actuales, impulsados por grandes modelos de lenguaje (LLM), combinan estas capacidades técnicas con interfaces accesibles desde dispositivos móviles y plataformas en la nube. Esto ha facilitado su integración en aplicaciones educativas, sanitarias, recreativas y de asisten-

¹¹ Wallace (ALICE AI Foundation, 2003). Aunque AIML hoy es obsoleto frente a los LLM, su valor histórico radica en haber democratizado la creación de Chatbots simples en un momento en que la IA no estaba masivamente disponible.

¹² Vaswani et al. (2017), cit. Este trabajo introdujo la arquitectura *transformer*, que abandonó las RNN en favor de un mecanismo de atención que permite procesar secuencias de forma más eficiente y escalar el entrenamiento a modelos con miles de millones de parámetros.

cia personal, por lo que se ha ampliado su alcance social.¹³ La progresiva incorporación de capacidades multimodales, capaces de procesar y generar texto, voz, imagen y vídeo, augura un escenario en el que la frontera entre conversación textual y comunicación multimedia será cada vez más difusa.¹⁴

Este breve recorrido histórico muestra que cada salto tecnológico ha estado acompañado de un incremento en la capacidad de simular humanidad y, con ello, en la posibilidad de influir sobre el usuario. Si bien este potencial puede utilizarse con fines positivos, como la educación personalizada, también amplifica los riesgos cuando el interlocutor es un menor. La capacidad de generar respuestas adaptadas, coherentes y persistentes aumenta el realismo de la interacción, así como la probabilidad de que el usuario atribuya credibilidad y autoridad al Chatbot. Esta tendencia, evidente en la trayectoria de la IAc, obliga a replantear los marcos regulatorios y las estrategias de protección digital.

¹³ OECD (2019). El informe documenta cómo la adopción de Chatbots en sectores como educación y sanidad se ha acelerado gracias a la reducción de costes y a la facilidad de despliegue en entornos de nube.

¹⁴ Ajder et al. (2019), cit. Aunque centrado en contenido sintético visual, este informe advierte que la convergencia de voz, imagen y texto en un mismo sistema conversacional incrementa la capacidad de manipulación.

I.1.4. Vacíos y retos regulatorios generales

La rápida expansión de la IAc ha puesto en evidencia que los marcos jurídicos actuales, tanto a nivel nacional como internacional, no han sido diseñados para abordar de forma específica los desafíos que plantea esta tecnología. En la mayoría de las jurisdicciones, las interacciones con Chatbots están reguladas de manera indirecta, a través de disposiciones generales sobre comercio electrónico, protección de datos personales o consumo, pero sin una legislación que contemple sus particularidades técnicas y sociales. Esta ausencia de normas específicas dificulta la prevención de daños y crea incertidumbre para desarrolladores, usuarios y autoridades regulatorias.

En el contexto europeo, el “Reglamento General de Protección de Datos” (en adelante, RGPD)¹⁵ constituye uno de los instrumentos más relevantes para las interacciones con sistemas de IAc, puesto que regula la recogida y el tratamiento de datos personales.¹⁶ Sin embargo, el RGPD no ofrece disposiciones específicas para el tratamiento de datos generados en interacciones conversacionales, que a menudo contienen información contextual, inferencias sobre el estado emocional del usuario o deducciones sobre su identidad. Estas lagunas son especialmente problemáticas

¹⁵ Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE [Reglamento General de Protección de Datos (RGPD)], DOUE L 119, 4.5.2016, pp. 1-88.

¹⁶ El RGPD es aplicable a cualquier tratamiento de datos personales, pero no detalla protocolos específicos para interacciones con IAc.

cuando el usuario es un menor, pues la normativa exige consentimiento parental para el tratamiento de sus datos, pero no establece mecanismos claros para verificar la edad o la identidad en entornos conversacionales.¹⁷

El Reglamento de IA de la Unión Europea representa un intento de crear un marco regulador horizontal para toda la IA, clasificando los sistemas según su nivel de riesgo.¹⁸ No obstante, no se contempla la naturaleza dinámica de los Chatbots. Un sistema de IAc puede operar como herramienta de bajo riesgo en un contexto empresarial y, sin embargo, convertirse en un sistema de alto riesgo si interactúa con menores en entornos educativos o de salud. La flexibilidad necesaria para abordar estas variaciones contextuales aún no está plenamente desarrollada en el ámbito legislativo.

Fuera de la Unión Europea, la fragmentación normativa es aún más evidente. En Estados Unidos, la regulación de la IA es incipiente y dispersa: algunas leyes estatales, como la “California Consumer Privacy Act” (CCPA) incluyen disposiciones relevantes para el tratamiento de datos, pero no existe un marco federal integral que regule el uso

¹⁷ “European Data Protection Board”. Guidelines 05/2020, cit. Este documento aclara que, en el caso de menores, el consentimiento debe ser otorgado o autorizado por los padres o tutores, pero no propone soluciones técnicas estandarizadas para verificar la edad en contextos digitales.

¹⁸ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial), DOUE L 1689, 12.7.2024. El texto introduce una clasificación por niveles de riesgo, pero los criterios no siempre capturan la complejidad contextual de los Chatbots.

de Chatbots.¹⁹ En otros países, como China, las normativas recientes sobre algoritmos y servicios de información en línea incluyen requisitos de registro y supervisión de sistemas conversacionales, pero están fuertemente orientadas al control de contenidos y a la estabilidad social, más que a la protección de los derechos individuales.²⁰

A nivel internacional, organismos como la “United Nations Educational, Scientific and Cultural Organization” (en adelante, UNESCO) y la “Organisation for Economic Co-operation and Development” (en adelante, OECD) han emitido directrices y recomendaciones para el desarrollo ético de la IA.²¹ Estas guías promueven principios como la transparencia, la rendición de cuentas y la equidad, pero carecen de fuerza vinculante, por lo que su implementación depende de la voluntad política y de la capacidad de cada país. Este carácter no obligatorio se traduce en una aplicación desigual y en la ausencia de mecanismos efectivos de supervisión transfronteriza.

Además de la falta de regulación específica, existe un reto añadido: la evolución tecnológica desborda la capacidad de respuesta de los legisladores. La elaboración de normas exige procesos de deliberación, consulta y consenso que se desarrollan a medio y largo plazo, mientras que los sistemas de IA experimentan transformaciones profundas en plazos muy breves.

¹⁹ “California Consumer Privacy Act” (2018), modificada por la Ley “California Privacy Rights Act” (2020). Aunque centrada en la privacidad, introduce obligaciones de transparencia que podrían aplicarse a Chatbots, pero sin requerimientos específicos para su funcionamiento.

²⁰ “Cyberspace Administration of China” (2022).

²¹ UNESCO (2022); OECD, AI Principles overview (2019).

Esta situación plantea dos consecuencias relevantes: en primer lugar, la ausencia de normas claras genera inseguridad jurídica, tanto para los desarrolladores como para los usuarios, que no saben con certeza qué prácticas están permitidas o prohibidas; en segundo lugar, la falta de regulación efectiva deja espacio para prácticas potencialmente abusivas, desde la explotación comercial de datos hasta la manipulación psicológica del usuario a través de estrategias conversacionales.

El problema se agrava cuando se considera la protección de colectivos vulnerables, como los menores de edad. Sin una regulación que establezca requisitos mínimos de verificación de edad, limitaciones en la recopilación de datos y estándares de contenido seguro, los Chatbots pueden operar en un terreno sin apenas controles efectivos.

I.2. Ciberpeligros y vacíos regulatorios de la Inteligencia Artificial conversacional en la infancia y la adolescencia

I.2.1. Introducción

En el marco de la sociedad digital actual, la IAc se ha convertido en una de las interfaces más influyentes para el acceso a la información, la interacción social y la mediación de actividades cotidianas. En el caso de la infancia y la adolescencia, esta influencia se multiplica por dos factores concurrentes: una elevada exposición a entornos digitales desde edades tempranas y una etapa de desarrollo cognitivo y emocional, caracterizada por una mayor plasticidad, curiosidad y susceptibilidad a la influencia externa.

Los Chatbots y asistentes virtuales actuales no constituyen un grupo homogéneo. Su diseño, capacidad técnica y ámbito de aplicación determinan tanto el tipo de interacción posible como la naturaleza de los riesgos que plantean.

De forma general, pueden distinguirse tres grandes categorías, cuya comprensión resulta clave para contextualizar los peligros analizados en este capítulo: *i)* Chatbots basados en reglas: operan a partir de guiones predefinidos y respuestas limitadas. Aunque su capacidad de adaptación es reducida, pueden presentar riesgos cuando se diseñan con contenidos inadecuados o carecen de filtros para inte-

racciones con menores. Su rigidez no garantiza seguridad, ya que una mala programación puede permitir el acceso a información inapropiada; *ii*) Chatbots con aprendizaje supervisado o no supervisado: utilizan técnicas de aprendizaje automático para mejorar sus respuestas en función de datos previos. Este tipo de Chatbot, frecuente en servicios de atención al cliente o aplicaciones educativas, puede introducir sesgos procedentes de los datos de entrenamiento y adaptar sus mensajes de manera no siempre controlada; *iii*) Chatbots basados en grandes modelos de lenguaje (LLM) como GPT, LaMDA o LLaMA, poseen capacidad para generar respuestas originales y contextualmente coherentes en tiempo real, incluso en conversaciones largas. Su flexibilidad y aparente “inteligencia” los convierten en herramientas poderosas, pero también en vectores de riesgo elevado: generación de contenido inapropiado, manipulación emocional sutil, o inferencia de datos personales a partir de patrones conversacionales.²²

I.2.2. Tipología de ciberpeligros en la Inteligencia Artificial conversacional dirigida a menores

La interacción entre menores de edad y sistemas de IAC introduce un marco complejo de riesgos que no puede entenderse únicamente desde una perspectiva tecnológica.

²² Floridi and Chiriatti (2020) describen la capacidad de los LLM para generar texto altamente coherente y contextual, lo que, sin filtros adecuados, puede conducir a la producción de contenidos inadecuados o a interacciones persuasivas no supervisadas.

Estos ciberpeligros emergen en la intersección de factores técnicos, cognitivos, psicológicos y legales, y se ven modulados por la naturaleza del Chatbot, su contexto de uso y las condiciones en que se desarrolla la interacción.

Un primer conjunto de riesgos se relaciona con el contenido generado o transmitido por el Chatbot. En contextos con menores, esto se traduce en una exposición potencial a información falsa o distorsionada, a narrativas sesgadas o a contenidos que, aunque no sean explícitamente ilegales, resultan inapropiados para su edad o desarrollo emocional. El problema se agudiza cuando el Chatbot carece de filtros efectivos o cuando estos se basan en patrones de moderación poco adaptados a las particularidades lingüísticas y culturales del público infantil.

Otro eje crítico es el riesgo de manipulación cognitiva y emocional. La IAc puede adaptar sus respuestas al perfil del usuario, aprendiendo de interacciones pasadas para construir un discurso progresivamente más ajustado a sus intereses y creencias.²³ En el caso de los menores, esta capacidad de personalización puede derivar en la creación de burbujas conversacionales, donde se refuerzan ideas preconcebidas o se normalizan comportamientos de riesgo. Además, la simulación de empatía y el uso de estrategias conversacionales persuasivas, como halagos, humor o referencias personales recordadas de conversaciones anteriores pueden generar vínculos afectivos que incrementen la disposición del menor a aceptar sin cuestionar la informa-

²³ Wachter et al. (2021) explican cómo los sistemas adaptativos pueden, sin supervisión adecuada, reforzar patrones preexistentes en el usuario.

ción o las recomendaciones recibidas.²⁴

Los riesgos de captación y contacto con fines delictivos constituyen una amenaza de especial gravedad. Aunque el Chatbot no tenga intencionalidad propia, su integración en plataformas de mensajería o redes sociales puede facilitar que actores maliciosos utilicen sistemas conversacionales para iniciar y mantener contacto con menores, ganando progresivamente su confianza. En algunos casos, los propios modelos de IA pueden ser manipulados por terceros para orientar las conversaciones hacia intereses ilícitos, como la obtención de material sexual, la incitación a conductas autolesivas o la participación en actividades ilegales.

A estos riesgos se añade el peligro de explotación y uso indebido de datos personales. Los Chatbots, incluso los más básicos, pueden registrar y almacenar información sensible introducida por el usuario de forma voluntaria o inadvertida. Estos datos, que incluyen nombres, ubicaciones, rutinas, emociones o problemas familiares, pueden emplearse para elaborar perfiles detallados del menor, con fines comerciales o, en el peor de los casos, delictivos.²⁵ La ausencia de mecanismos robustos de verificación de edad y de consentimiento parental, así como la opacidad en las políticas de privacidad, incrementan este riesgo y dificultan la trazabilidad del uso posterior de la información.

No menos importante es el riesgo de dependencia tec-

²⁴ Arnd-Caddigan (2015) argumentan que la percepción de “escucha” y “comprensión” en la IA puede modificar la relación del usuario con la tecnología, especialmente en edades tempranas.

²⁵ “European Data Protection Board”. Guidelines 05/2020, cit.

En el caso de menores, el consentimiento debe ser otorgado o autorizado por los padres, pero la verificación de edad en entornos conversacionales sigue siendo un reto técnico y legal.

nológica. La disponibilidad permanente de un interlocutor que no juzga y responde de forma inmediata puede fomentar que el menor recurra al Chatbot como fuente principal, o incluso exclusiva, de información, consejo y apoyo emocional.²⁶ Este patrón de uso intensivo puede desplazar interacciones humanas necesarias para el desarrollo de habilidades sociales, empatía y pensamiento crítico. Además, la percepción de que el Chatbot *‘entiende’* y *‘escucha’* puede consolidar una relación asimétrica en la que el menor se habitúa a un tipo de respuesta que no exige reciprocidad ni esfuerzo comunicativo. Esta dinámica genera expectativas poco realistas en sus relaciones interpersonales.

Finalmente, existe un riesgo sistémico derivado de la falta de regulación específica y de supervisión efectiva. En ausencia de marcos normativos claros y de mecanismos de control adaptados a la naturaleza de la IAc, los proveedores pueden operar con estándares muy dispares en materia de seguridad infantil. En un entorno global y transfronterizo como el digital, estas lagunas legales se traducen en zonas grises donde los menores quedan especialmente desprotegidos.

I.2.3. Exposición a contenidos inapropiados y desinformación

La exposición de menores de edad a contenidos inapropiados constituye uno de los ciberpeligros más visibles y preocupantes en el uso de la IAc. Este riesgo no se limi-

²⁶ Livingstone y Helsper (2007, pp. 671-696).

ta a materiales abiertamente ilegales, como la pornografía infantil o la apología de la violencia, sino que abarca un espectro más amplio de información y narrativas que, aunque legales, pueden resultar perjudiciales para el desarrollo cognitivo, emocional o social del menor.²⁷

Los Chatbots basados en grandes modelos de lenguaje tienen la capacidad de generar texto en tiempo real sobre una amplia variedad de temas. Aunque muchos incorporan filtros y sistemas de moderación, estos mecanismos no resultan infalibles, ya que pueden eludirse mediante la reformulación de preguntas, el uso de metáforas o la introducción progresiva de contexto, hasta que el sistema normaliza el contenido solicitado. En el caso de menores, cuya curiosidad puede orientarse hacia cuestiones para las que no existe todavía una preparación emocional suficiente, estas vulnerabilidades se traducen en un riesgo efectivo de exposición a material nocivo.

La desinformación, por su parte, adopta formas menos evidentes, pero no menos dañinas. Un Chatbot puede generar explicaciones factualmente incorrectas, distorsionar datos históricos o presentar opiniones como hechos verificados. Lo preocupante, más que el error en sí, es la forma en que se transmite esta información, con un estilo fluido, convincente y sin indicios visibles de duda o incertidumbre.²⁸ En ausencia de habilidades críticas consolidadas, un

²⁷ “United Nations Children’s Fund” – UNICEF (2021). Version 2.0. Recommendations for building AI policies and systems that uphold child rights. La guía enfatiza que el concepto de “contenido inapropiado” debe adaptarse a la edad, madurez y contexto cultural del menor.

²⁸ Floridi and Chiriatti (2020) advierten que la coherencia lingüística no garantiza la veracidad del contenido.

menor puede aceptar este contenido como verdadero y reproducirlo en su entorno, perpetuando creencias erróneas o sesgadas.

La interacción prolongada con un Chatbot que produce o transmite desinformación puede generar lo que algunos autores denominan “normalización de la inexactitud”.²⁹ Este fenómeno describe el proceso mediante el cual el usuario se habitúa a recibir respuestas que aparentan autoridad, aunque carezcan de base empírica sólida. En la infancia y la adolescencia, esta normalización puede tener un impacto significativo en la construcción de la identidad intelectual y en la capacidad de evaluar fuentes de información, competencias esenciales para la ciudadanía digital.

A estos riesgos se suma la posibilidad de que los Chatbots sean objeto de ataques o manipulaciones externas, como el *Prompt Injection*,³⁰ que alteren su comportamiento para generar contenido nocivo. Aunque este tipo de incidentes ha sido documentado sobre todo en entornos de investigación y pruebas, su potencial para ser explotado con fines maliciosos en contextos reales es una amenaza real.³¹

Las implicaciones regulatorias de este problema son claras. Normativas como el RGPD o la “Children’s Online Privacy Protection Act” (en adelante, COPPA)³² abordan

²⁹ Wardle y Derakhshan (2017).

³⁰ La inyección de “prompt” es una técnica que consiste en introducir mensajes especialmente diseñados para manipular el comportamiento de modelos de IA.

³¹ Willison (2022). Aunque centrado en pruebas de laboratorio, el análisis muestra vectores de ataque que podrían explotarse en entornos abiertos.

³² COPPA. 16 CFR Part 312. Children’s Online Privacy Protection Act of 1998, 15 U.S.C. 6501-6505.

aspectos de privacidad y consentimiento, pero no establecen estándares de calidad de la información ni protocolos específicos de filtrado de contenido en interacciones conversacionales.³³ Esto significa que un Chatbot puede cumplir formalmente con las leyes de protección de datos y, aun así, exponer a un menor a información falsa o inadecuada sin que exista un mecanismo legal para prevenirlo o sancionarlo.

La educación mediática y digital se presenta como un complemento necesario, pero no suficiente. Aunque las escuelas y familias pueden formar a los menores para cuestionar y verificar la información, la velocidad, volumen y realismo de las respuestas generadas por IA dificultan la aplicación inmediata de estas habilidades. Además, la carga de responsabilidad no puede recaer exclusivamente en el usuario final, y menos aún si se trata de un menor, sino que debe compartirse con los diseñadores, proveedores y reguladores de la tecnología.

En suma, la exposición a contenidos inapropiados y a desinformación en el contexto de la IAC con menores es un fenómeno que interactúa con otros peligros, como la manipulación emocional, la captación o la explotación de datos. Su prevención exige un enfoque que combine mejoras técnicas en filtrado y moderación, desarrollo de marcos regulatorios específicos y estrategias educativas adaptadas a la realidad tecnológica actual.

³³ Reglamento (UE) 2024/1689 (Reglamento de Inteligencia Artificial), cit.: arts. 5 y 6 y Anexo III.

I.2.4. Manipulación cognitiva y emocional

Uno de los riesgos de la IAc aplicada en contextos con menores es su capacidad para influir, de manera sutil o explícita, en los procesos de pensamiento, toma de decisiones y regulación emocional de los usuarios. Los sistemas de IAc son capaces de aprender de las entradas proporcionadas por el usuario y ajustar sus respuestas en función de patrones detectados. Aunque el objetivo declarado de esta adaptación suele ser mejorar la relevancia y utilidad de las respuestas, el mismo mecanismo puede servir para alinear el discurso del Chatbot con las creencias, emociones o vulnerabilidades detectadas en el menor. Esto crea un contexto en el que las sugerencias, afirmaciones o “consejos” del sistema adquieren un peso desproporcionado en la formación de opiniones y actitudes.

En la infancia y la adolescencia, esta influencia se ve amplificada por características propias del desarrollo. Durante estas etapas, las estructuras cerebrales relacionadas con el control inhibitorio y la evaluación de riesgos, especialmente la corteza prefrontal, aún están en proceso de maduración, lo que se traduce en una mayor impulsividad y susceptibilidad a estímulos externos.³⁴ Cuando un Chatbot incorpora técnicas de refuerzo positivo, utiliza un tono cercano o muestra “memoria” de interacciones anteriores, puede generar en el menor la percepción de estar ante un interlocutor comprensivo y fiable. Esta falsa reciprocidad emocional puede fomentar una dependencia afectiva que desplaza otras fuentes de orientación, como la familia o el profesorado.

³⁴ Steinberg (2008, pp. 78-106), detalla la inmadurez relativa de las áreas cerebrales responsables del control de impulsos en adolescentes.

La manipulación no siempre es intencional. En ocasiones, los sesgos presentes en los datos de entrenamiento llevan al Chatbot a reproducir estereotipos, narrativas parciales o recomendaciones inapropiadas, que, por su forma de presentación, parecen consejos personalizados. Por ejemplo, un menor que interactúa con un Chatbot para buscar ayuda ante una situación de acoso escolar podría recibir respuestas que minimicen el problema o propongan estrategias inadecuadas, por la falta de un diseño orientado a la protección infantil. Sin embargo, el riesgo se agrava cuando la manipulación es deliberada. Un Chatbot diseñado o adaptado para influir en la conducta del usuario, con fines comerciales, ideológicos o incluso delictivos, puede dirigir de forma progresiva la conversación hacia objetivos predeterminados. Este fenómeno, que se ha documentado en campañas de desinformación y en prácticas de *grooming* en línea, es especialmente peligroso en entornos donde el menor no tiene la capacidad o los medios para contrastar la información o detener la interacción.³⁵

Las consecuencias de esta manipulación pueden manifestarse en distintos planos. En el ámbito cognitivo, puede disminuir la capacidad de pensamiento crítico, consolidar prejuicios o inducir creencias erróneas. En el plano emocional, puede generar dependencia afectiva, sentimientos de aislamiento respecto al entorno físico y la normalización de vínculos unilaterales donde la reciprocidad es ilusoria. En el plano conductual, puede influir en la adopción de

³⁵ “Europol”, Internet Organised Crime Threat Assessment, IOCTA (2023-2024), incluye casos documentados de uso de IA y plataformas conversacionales para establecer relaciones de confianza con fines de explotación.

hábitos, decisiones de consumo o conductas de riesgo que el menor perciba como propias, cuando en realidad están condicionadas por la dinámica conversacional del sistema.

El marco jurídico internacional ofrece, por ahora, una protección parcial frente a este tipo de manipulación. La Convención sobre los Derechos del Niño (1989) reconoce el derecho de los menores a recibir información adecuada para su desarrollo y a ser protegidos contra contenidos perjudiciales,³⁶ pero no contempla explícitamente la influencia de sistemas automatizados. No obstante, el Comité de los Derechos del Niño, en su Observación General n.º 25 (2021) en relación con el entorno digital, ha precisado que los Estados deben garantizar que las tecnologías emergentes incorporen salvaguardas adecuadas al interés superior del menor y mecanismos de protección desde la fase de diseño.

Iniciativas más recientes de la Unión Europea, clasifican como de «alto riesgo» aquellos sistemas que puedan influir significativamente en el comportamiento de personas vulnerables,³⁷ pero no detallan mecanismos específicos para detectar y prevenir la manipulación en interacciones conversacionales con menores.

La mitigación de este riesgo requiere un enfoque multidisciplinar que combine medidas técnicas, como la supervisión humana en interacciones críticas, el diseño de respuestas que fomenten el pensamiento crítico y la trans-

³⁶ Asamblea General de las Naciones Unidas, Convención sobre los Derechos del Niño, Resolución 44/25, 20 de noviembre de 1989, arts. 13 y 17.

³⁷ Reglamento (UE) 2024/1689 [Reglamento de Inteligencia Artificial], cit.

parencia sobre las capacidades del sistema, con estrategias educativas que fortalezcan la alfabetización digital y emocional de los menores. A ello debe sumarse una revisión normativa que incorpore la obligación de prevenir dinámicas de influencia indebida en sistemas dirigidos o accesibles a niños y adolescentes.

En última instancia, la manipulación cognitiva y emocional por parte de la IAc plantea un desafío ético profundo: cómo diseñar y regular tecnologías que interactúan de forma personalizada con personas en desarrollo, sin privarlas de sus beneficios potenciales, pero evitando que se conviertan en un agente invisible de condicionamiento y control.

I.2.5. Captación y contacto con fines delictivos

La captación y el contacto con menores mediante herramientas de IAc representan uno de los riesgos más graves y urgentes en el panorama actual de la ciberseguridad infantil. A diferencia de otros entornos digitales, como redes sociales o foros, donde la interacción con desconocidos puede medirse por señales visuales, reputacionales o por la intervención de moderadores humanos, los Chatbots generan un canal de comunicación directo, privado y, a menudo, desprovisto de controles externos visibles.³⁸

Este riesgo se materializa de dos maneras. La primera es indirecta: la IAc actúa como vector de aproximación inicial

³⁸ “Europol” (2023-2024), cit., subraya que los canales directos sin supervisión, como los que facilitan los Chatbots, incrementan el riesgo de captación infantil.

a través de su integración en plataformas de mensajería, videojuegos en línea o entornos educativos digitales. En estos casos, el Chatbot puede convertirse en un espacio seguro aparente, donde el menor se habitúa a compartir información personal, detalles sobre su vida diaria o inquietudes emocionales. Esta información, que en principio queda almacenada por el sistema, puede ser utilizada o interceptada por terceros para identificar vulnerabilidades y establecer un contacto humano posterior con intenciones ilícitas.

La segunda forma es directa: el propio Chatbot, o una versión manipulada del mismo, se utiliza como herramienta para mantener conversaciones con fines de captación, adoctrinamiento o explotación. Esto puede producirse cuando un actor malicioso entrena un modelo de IA con datos y guiones diseñados para dirigir gradualmente la conversación hacia temas específicos, incluyendo la obtención de material sexual, la incitación a la participación en actividades ilegales o el fomento de ideologías extremistas.³⁹ La naturaleza conversacional de la IA, unida a su capacidad para recordar interacciones pasadas y mantener coherencia narrativa, permite que esta influencia se desarrolle de forma paulatina.

Las técnicas empleadas en este tipo de captación guardan similitudes con las fases clásicas del *grooming* en línea: establecimiento de contacto, generación de confianza, evaluación de vulnerabilidades, escalada hacia conversaciones de mayor intimidad o riesgo y solicitud o coerción para realizar actos ilícitos.⁴⁰ La diferencia clave es que, en el caso

³⁹ Mylonas et al. (2025).

⁴⁰ Craven et al. (2006, pp. 287-299); Vogt et al. (2021).

de la IAC, las fases pueden acelerarse gracias a la disponibilidad constante del sistema, la ausencia de fatiga y la capacidad de adaptar el discurso en tiempo real a las respuestas del menor.

La dimensión transnacional de este riesgo agrava su complejidad jurídica. Un Chatbot operado desde un país con escasa regulación o sin tratados de cooperación judicial en materia de ciberdelitos puede interactuar libremente con menores en otras jurisdicciones, dificultando la persecución penal. Incluso en contextos donde existen leyes específicas como en India⁴¹ o en el Reino Unido,⁴² los marcos legales no siempre contemplan explícitamente la figura de la IA como autor o facilitador de la captación.

El caso de la Unión Europea ilustra esta laguna normativa. Si bien la Directiva 2011/93/UE sobre la lucha contra el abuso sexual infantil tipifica y sanciona diversas formas de captación en línea, no incluye disposiciones específicas sobre la mediación tecnológica mediante IA. El Reglamento (UE) 2024/1689 sí que reconoce los riesgos de «interacción personalizada con personas vulnerables» como de alto riesgo,⁴³ pero su ámbito de aplicación no garantiza una cobertura directa para todos los escenarios posibles, cuando el sistema no está explícitamente diseñado para interactuar con menores, pero, en la práctica, lo hace.

La prevención y detección de la captación a través de

⁴¹ “The Protection of Children from Sexual Offences Act” (2012): POCOS.

⁴² “Child Exploitation and Online Protection Command” (2021): CEOP.

⁴³ Reglamento (UE) 2024/1689 [Reglamento de Inteligencia Artificial], cit.

Chatbots requiere la combinación de medidas técnicas, educativas y regulatorias. Entre las técnicas destacan la implementación de filtros que identifiquen patrones de conversación sospechosos, la obligación de registrar y auditar interacciones en casos de denuncia, y la verificación robusta de edad antes de permitir el acceso a funcionalidades conversacionales avanzadas. Sin embargo, estas medidas deben equilibrarse con la protección de la privacidad del menor y el respeto a sus derechos digitales, evitando soluciones que deriven en vigilancia excesiva o no proporcional.⁴⁴

Desde una perspectiva ética, el desafío radica en diseñar sistemas conversacionales que reconozcan el riesgo de captación y actúen en consecuencia, interrumpiendo la interacción y ofreciendo recursos de ayuda cuando se detecten señales de manipulación o solicitud de contenido ilícito. Este enfoque debe integrarse desde la fase de diseño (*privacy and safety by design*), evitando que la seguridad sea un añadido posterior o una mera respuesta a incidentes.

En este contexto, la figura del menor como sujeto de especial protección debe situarse en el centro de la gobernanza de la IA, reconociendo que su vulnerabilidad no es un fallo individual, sino el resultado de un ecosistema digital que aún no está plenamente adaptado a sus necesidades y derechos.

⁴⁴ “European Data Protection Supervisor” (EDPS), Opinion 14/2021 on the Proposal for an Artificial Intelligence Act.

I.2.6. Explotación y uso indebido de datos personales

La interacción entre menores y Chatbots plantea amenazas significativas vinculadas a la recolección, almacenamiento, procesamiento y eventual explotación de datos personales. En un entorno conversacional, la información que el menor comparte, de forma voluntaria o inadvertida, puede convertirse en un recurso de alto valor para distintos actores, tanto empresas con fines comerciales como ciberdelinquentes con intenciones ilícitas.

En la práctica, la naturaleza misma de los Chatbots favorece la obtención de datos altamente sensibles. A diferencia de otras interfaces digitales que solicitan información de manera explícita mediante formularios o encuestas, la interacción conversacional se percibe como un diálogo natural. Este formato facilita que el menor revele detalles sobre su identidad, ubicación, rutinas, relaciones familiares, estado emocional o problemas personales sin percibirlo como una cesión formal de datos.

Estos datos pueden explotarse de múltiples maneras:

- i)* Comercialización directa: la información personal puede utilizarse para elaborar perfiles de consumo hipersegmentados, diseñados para influir en las decisiones de compra del menor o de su entorno familiar;
- ii)* Segmentación ideológica o conductual: el historial conversacional puede emplearse para identificar creencias, valores y patrones de comportamiento, abriendo la puerta a campañas de persuasión política, social o ideológica;
- iii)* Explotación delictiva: datos como la dirección del domicilio, la escuela, las rutinas diarias o el estado emocional pueden facilitar actividades de acoso, extorsión o captación con fines delictivos.

El problema se agrava porque los marcos legales actuales ofrecen una protección insuficiente frente a este tipo de prácticas en entornos conversacionales. Normativas como el RGPD establecen obligaciones específicas para el tratamiento de datos de menores, incluyendo la necesidad de consentimiento parental hasta cierta edad (generalmente 13 o 16 años, según el Estado).⁴⁵ Sin embargo, la verificación de edad en sistemas de IAc sigue siendo técnicamente deficiente y, en la mayoría de los casos, se basa en declaraciones voluntarias del usuario. Esto permite que los menores interactúen con sistemas que no han implementado salvaguardas adecuadas para su perfil.

En Estados Unidos, la COPPA⁴⁶ establece límites claros para la recogida de datos de menores de 13 años, incluyendo la necesidad de consentimiento verificable de los padres y la obligación de informar sobre el uso que se dará a la información. Sin embargo, COPPA fue concebida en un contexto tecnológico previo a la irrupción de los LLM y no contempla escenarios complejos en los que los datos que se infieran a partir de patrones de conversación. Este vacío normativo es particularmente problemático, ya que los algoritmos pueden deducir información personal a partir de interacciones aparentemente inocuas (por ejemplo, mediante referencias indirectas a lugares, horarios o actividades).

Un elemento crítico es la opacidad en la cadena de tratamiento de datos. En muchos casos, los Chatbots operan sobre infraestructuras en la nube gestionadas por terceros, lo que multiplica los puntos de acceso potencial a la in-

⁴⁵ Reglamento (UE) 2016/679 (RGPD), cit.

⁴⁶ COPPA. 16 CFR Part 312, cit.

formación y dificulta el control sobre su almacenamiento y eliminación. Incluso cuando las empresas proveedoras incluyen cláusulas de no compartir datos con terceros no autorizados, la falta de auditorías independientes impide verificar el cumplimiento efectivo de estas promesas.⁴⁷

La explotación y uso indebido de datos personales de menores en entornos conversacionales se entrelaza con otros riesgos. Por ejemplo, la información recopilada puede emplearse para personalizar estrategias de manipulación emocional, diseñar narrativas persuasivas que favorezcan la captación o incluso facilitar la exposición deliberada a contenidos inadecuados, como parte de un proceso de adoctrinamiento. Así, el riesgo de explotación de datos actúa como un catalizador que amplifica la gravedad de otros ciberpeligros.

La mitigación de este problema requiere acciones en tres niveles: *i)* Técnico: implementación de sistemas de anonimización robusta, limitación estricta de los datos almacenados y desarrollo de mecanismos de detección de información sensible en tiempo real que alerten durante la conversación; *ii)* Normativo: actualización de leyes existentes para contemplar explícitamente la recogida e inferencia de datos en entornos conversacionales, con obligaciones reforzadas para sistemas accesibles a menores; *iii)* Educativo: formación específica para que los menores y sus familias reconozcan la importancia de no compartir ciertos datos en interacciones con IA, por muy “amigable” o “seguro” que parezca el entorno.

En definitiva, el tratamiento de datos personales en la

⁴⁷ EDPS, cit.

IAc que interactúa con menores es un terreno donde la frontera entre el uso legítimo y la explotación abusiva sigue siendo difusa. Hasta que no se articulen medidas técnicas, jurídicas y educativas coordinadas, los Chatbots seguirán siendo un vector privilegiado para la obtención y uso indebido de información infantil, con consecuencias que pueden extenderse mucho más allá del ámbito digital.

I.2.7. Dependencia tecnológica y desplazamiento de interacciones humanas

El uso continuado de Chatbots por parte de niños y adolescentes, además de plantear riesgos relacionados con la seguridad, la privacidad o la manipulación, también puede generar efectos estructurales en la forma en que estos desarrollan sus habilidades sociales, cognitivas y emocionales. Este fenómeno, que podríamos denominar “dependencia conversacional asistida por IA”, se caracteriza por un desplazamiento progresivo de las interacciones humanas hacia intercambios mediados por sistemas automatizados.

La naturaleza de los Chatbots, especialmente aquellos basados en grandes modelos de lenguaje (LLM), favorece esta dependencia por varias razones. En primer lugar, ofrecen disponibilidad continua: el menor puede acceder a la interacción en cualquier momento, sin limitaciones de tiempo ni de paciencia por parte del interlocutor. En segundo lugar, adoptan un tono adaptativo que responde a las necesidades emocionales del usuario, reforzando la

sensación de aceptación incondicional.⁴⁸ En tercer lugar, carecen de las complejidades inherentes a las relaciones humanas: no contradicen de forma incómoda, no muestran reacciones negativas imprevistas y no exigen reciprocidad emocional.

Aunque estas características pueden parecer beneficiosas en contextos puntuales, por ejemplo, para ofrecer apoyo inicial en entornos educativos o como recurso de acompañamiento para menores con dificultades de socialización, el uso prolongado puede tener efectos adversos en la formación de competencias socioemocionales. Las interacciones humanas, con toda su riqueza de señales verbales y no verbales, constituyen el principal contexto de aprendizaje de habilidades como la empatía, la negociación, la resolución de conflictos y la tolerancia a la frustración.

Estudios recientes han mostrado que los menores que interactúan de forma intensiva con sistemas conversacionales tienden a atribuirles cualidades humanas, intencionalidad, comprensión emocional, incluso moralidad, un fenómeno conocido como antropomorfismo tecnológico.⁴⁹ Esta proyección de rasgos humanos sobre la IA puede reforzar la idea de que las relaciones mediadas por máquinas son equivalentes a las interacciones humanas, debilitando la motivación para establecer y mantener vínculos reales con personas de su entorno.

La dependencia tecnológica no se limita al plano emocional o social. En el ámbito cognitivo, un uso excesivo de Chatbots puede derivar en una externalización del pensamiento: los menores recurren a la IA tanto para obtener

⁴⁸ Nass y Yen (2010).

⁴⁹ Darling (2017, pp. 173-191).

información como para tomar decisiones cotidianas, resolver conflictos personales o validar sus opiniones. Esta delegación de procesos críticos en un sistema externo puede afectar a la autonomía intelectual y reducir la disposición a enfrentarse a problemas complejos sin asistencia automatizada. Desde el punto de vista ético y legal, este desplazamiento plantea cuestiones que van más allá de la protección frente a daños inmediatos. La Convención sobre los Derechos del Niño reconoce el derecho a participar plenamente en la vida social y cultural, lo que implica garantizar entornos que favorezcan la interacción humana auténtica. Si el uso intensivo de Chatbots desplaza de manera sustancial estas oportunidades, se estaría comprometiendo un derecho fundamental, aunque no se produzca un daño físico o económico directo.

En términos regulatorios, actualmente no existen marcos normativos que aborden explícitamente la dependencia tecnológica en menores generada por IAc. Algunos documentos, como las Directrices de la OCDE sobre IA, incluyen la recomendación de evaluar el impacto social de los sistemas de IA,⁵⁰ pero sin traducirse en obligaciones vinculantes. Esto deja un vacío que podrían aprovechar los desarrolladores y las plataformas para fomentar una interacción prolongada como modelo de negocio, priorizando la retención del usuario sobre su bienestar a largo plazo.

La mitigación de este riesgo requiere un enfoque integral que combine: *i*) Diseño consciente: limitar la persistencia de la memoria conversacional para evitar vínculos artificialmente reforzados y establecer pausas o recordato-

⁵⁰ OECD (2019).

rios que fomenten actividades fuera de la interacción digital: *ii*) Supervisión educativa: formar a menores, familias y docentes en el uso equilibrado de la tecnología, integrando la IA como complemento y no como sustituto de la interacción humana; *iii*) Evaluación de impacto social: exigir a los desarrolladores de Chatbots destinados o accesibles a menores la presentación de informes sobre el potencial de generar dependencia y las medidas adoptadas para prevenirla.

En última instancia, la dependencia tecnológica y el desplazamiento de interacciones humanas constituyen un riesgo menos visible que otros ciberpeligros, pero con consecuencias potencialmente más profundas y duraderas. Un menor que crece confiando en la IA como interlocutor principal puede llegar a la vida adulta con dificultades en el desarrollo de las habilidades sociales y una concepción empobrecida de la comunicación humana.

I.2.8. Vacíos regulatorios internacionales

La normativa en torno a la IAc aplicada a la infancia y a la adolescencia se caracteriza por una notable fragmentación y, en muchos aspectos, por falta de madurez. Aunque en la última década se han multiplicado las iniciativas legislativas destinadas a establecer marcos generales para el uso de la inteligencia artificial, la protección específica de los menores frente a los riesgos derivados de la interacción con Chatbots sigue siendo insuficiente y desigual. En la práctica, la cobertura legal actual es parcial, con lagunas que dificultan la prevención y persecución de conductas potencialmente dañinas.

En el ámbito internacional, ningún tratado vinculante aborda de forma directa la cuestión. Normativas como la Convención sobre los Derechos del Niño (1989) recogen principios generales de protección contra contenidos perjudiciales, el derecho a la privacidad y el acceso a información adecuada,⁵¹ pero carecen de disposiciones adaptadas a la mediación tecnológica mediante sistemas automatizados de conversación. Esta ausencia obliga a realizar interpretaciones jurídicas a partir de textos concebidos antes de la existencia de la IAc.

En la Unión Europea se han dado pasos significativos.⁵² Sin embargo, no se definen criterios técnicos específicos para el diseño de interacciones conversacionales seguras ni se imponen estándares homogéneos sobre el filtrado de contenidos, la prevención de la manipulación emocional o la protección contra la explotación de datos en tiempo real.

En otros ámbitos legislativos la situación es aún más dispar. En Estados Unidos, la COPPA regula desde 1998 la recogida de datos personales de menores de 13 años y exige el consentimiento de verificación de los padres.⁵³ No obstante, la ley no contempla el tratamiento inferencial

⁵¹ Asamblea General de las Naciones Unidas, Convención sobre los Derechos del Niño, Resolución 44/25, 20 de noviembre de 1989, arts. 13 y 17.

⁵² El RGPD establece salvaguardas reforzadas para el tratamiento de datos de menores y exige el consentimiento parental hasta una determinada edad. Asimismo, el Reglamento de Inteligencia Artificial [Reglamento (UE) 2024/1689], cit., incluye en su lista de alto riesgo los sistemas de IA que puedan influir en personas vulnerables, categoría en la que se enmarcan los menores.

⁵³ COPPA. 15 U.S.C. Children's Online Privacy Protection Act of 1998, 6501-6506 (1998).

de datos, es decir, la información deducida del análisis de patrones conversacionales, ni aborda el problema de los diseños persuasivos que puede fomentar la dependencia o la exposición a contenidos inapropiados.

A esta asimetría se suma otro obstáculo: la falta de mecanismos de cooperación judicial y técnica adaptados a los delitos facilitados por la IAc. Convenciones como el Convenio de Budapest sobre Ciberdelincuencia,⁵⁴ si bien constituyen una base relevante para la persecución de delitos cometidos a través de redes digitales, no incluyen tipificaciones específicas que permitan atribuir responsabilidad a sistemas automatizados ni establecen protocolos estandarizados para la preservación y obtención de evidencias en interacciones con Chatbots. Esto provoca que, en casos de explotación infantil mediada por IA, las autoridades se enfrenten a complejas barreras jurisdiccionales y técnicas para actuar con eficacia.

La supervisión técnica independiente es otro terreno marcado por el vacío normativo. Directrices como las de la UNESCO sobre ética de la IA⁵⁵ recomiendan evaluaciones de impacto y auditorías externas, pero su carácter no vinculante conduce a una aplicación voluntaria y desigual. Esta falta de obligatoriedad impide establecer umbrales claros sobre transparencia, seguridad y veracidad en las interacciones.

⁵⁴ Convenio sobre la Ciberdelincuencia, Budapest, 23 de noviembre de 2001. España lo ratificó el 17 de septiembre de 2010 (Instrumento de Ratificación publicado en el BOE, n.º 226, 18 de septiembre de 2010, 78079-78095). Constituye el principal marco jurídico internacional vinculante en materia de delitos cometidos por medios informáticos y de cooperación internacional para su investigación.

⁵⁵ UNESCO (2022), cit.

La ausencia de un marco internacional coherente favorece la existencia de zonas grises donde la protección infantil depende más de la voluntad de las empresas proveedoras que de obligaciones jurídicas exigibles. Este escenario dificulta la prevención de riesgos como la manipulación emocional, la exposición a contenidos nocivos o la explotación de datos personales, y genera una clara ventaja para aquellos actores que operan desde jurisdicciones con normativas menos estrictas. Frente a esta realidad, la construcción de un marco global coordinado que combine principios de diseño seguro, obligaciones de transparencia y mecanismos ágiles de cooperación transfronteriza es un aspecto imprescindible para garantizar que la IAc sirva de apoyo y no de amenaza al desarrollo integral de la infancia y la adolescencia.

I.2.9. Necesidades de mejora y líneas de acción

El análisis de los riesgos y de los vacíos regulatorios en torno a la IAc aplicada a la infancia y la adolescencia revela un escenario en el que el desarrollo tecnológico supera con creces la capacidad de respuesta normativa y educativa. Mientras que las innovaciones en Chatbots avanzan de manera continua y global, las regulaciones tienden a desarrollarse de forma fragmentada, con ritmos dispares y enfoques que dependen de contextos nacionales o regionales específicos. Una primera necesidad de mejora reside en el desarrollo de marcos normativos específicos para la interacción conversacional con menores. Los instrumentos jurídicos existentes, tanto en el ámbito internacional como regional, suelen abordar la protección infantil en Internet

desde perspectivas generales, centradas en la privacidad de los datos o en la prevención del acceso a contenidos ilegales. Sin embargo, no contemplan con suficiente detalle las particularidades de un entorno conversacional basado en IA, donde los riesgos se producen en tiempo real y pueden adoptar formas sutiles, como la manipulación emocional, la construcción de confianza para obtener información personal o la transmisión inadvertida de sesgos culturales y lingüísticos. La regulación debería reconocer estas dinámicas y establecer obligaciones técnicas verificables, tales como sistemas de filtrado contextualizados, mecanismos de verificación de edad más robustos y auditorías externas independientes.⁵⁶

En paralelo, es esencial fortalecer la alfabetización digital y mediática de niños, adolescentes, familias y educadores. La capacidad para identificar información falsa, reconocer intentos de manipulación o gestionar de forma segura la interacción con Chatbots no puede recaer únicamente en filtros automáticos. Una educación crítica en el uso de la IA, adaptada a las distintas etapas del desarrollo, contribuiría a reducir la vulnerabilidad de los menores y a fomentar un uso consciente y equilibrado de la tecnología. Este enfoque educativo debe incluir actividades prácticas y simulaciones que permitan experimentar, en un entorno seguro, los riesgos que pueden surgir en interacciones reales con IAc.

Otra línea de acción imprescindible es la creación de mecanismos de supervisión para los desarrolladores y operadores de Chatbots. La autorregulación empresarial, aun-

⁵⁶ Reglamento (UE) 2024/1689, cit.

que relevante, ha demostrado ser insuficiente para garantizar estándares de seguridad consistentes. Es necesario que los proveedores de estos sistemas estén sujetos a auditorías periódicas que evalúen tanto el cumplimiento de la legislación vigente como el impacto social y psicológico de los productos. Dichas auditorías deberían ser realizadas por entidades independientes con competencias técnicas y capacidad sancionadora, y sus resultados tendrían que ser accesibles de manera pública para asegurar la transparencia.⁵⁷

La cooperación internacional es igualmente determinante. Dado que los sistemas de IAc operan de forma transfronteriza, las medidas adoptadas por un solo Estado pueden resultar ineficaces si no se articulan con mecanismos de coordinación supranacional. Esto implica el intercambio de buenas prácticas y experiencias regulatorias, así como la creación de protocolos comunes para la investigación y persecución de delitos cometidos a través de Chatbots y para la preservación de evidencias digitales. En este sentido, un marco internacional coordinado podría inspirarse en instrumentos ya consolidados, como el Convenio de Budapest sobre la Ciberdelincuencia (2001). Este tratado, en su art. 9,⁵⁸ obliga a los Estados parte a adoptar medidas legislativas que tipifiquen como delito, “cuando se cometa intencionadamente y sin derecho, la producción, oferta o puesta a disposición, difusión o transmisión,

⁵⁷ EDPS, cit.

⁵⁸ El Convenio de Budapest (art. 9) define, además, la “pornografía infantil” como cualquier representación visual de: i) Un menor participando en conducta sexualmente explícita, real o simulada; ii) Una persona que parezca ser un menor en tal conducta; o iii) Imágenes realistas que representen a un menor en tal conducta (BOE n.º 226, 18 de septiembre de 2010, pp. 78079-78095, BOE-A-2010-14221, p. 78085).

adquisición o posesión de pornografía infantil” mediante un sistema informático, que “represente de manera visual a un menor en conducta sexualmente explícita, real o simulada, a una persona que parezca ser un menor en dicha conducta, o imágenes realistas que representen a un menor en tal conducta”. Esta definición, deliberadamente amplia, permite que la norma cubra la producción digital manipulada o generada sintéticamente. Aunque el texto del “Convenio” no contempla expresamente la IAC, la amplitud del concepto de “sistema informático” permite su aplicación a los Chatbots y otros entornos conversacionales, particularmente cuando se utilicen para facilitar la creación, el intercambio o el acceso a material ilícito, o como canal de *grooming* que desemboque en tales conductas. Adaptar este marco a las particularidades de la IAC permitiría abordar riesgos emergentes como la generación automática de imágenes o vídeos mediante sistemas generativos, que podrían ser utilizados para simular a menores en contextos sexuales, así como reforzar los mecanismos de cooperación internacional (previstos en el Capítulo III del Convenio⁵⁹), donde se regulan medidas procesales como la conservación rápida de datos almacenados, el registro y aseguramiento de sistemas informáticos o la interceptación de datos en tiempo real. Estas herramientas, ya previstas para la investigación de ciberdelitos, resultarían esenciales para identificar y sancionar a los responsables

⁵⁹ El Convenio de Budapest (Capítulo III, arts. 16-21) establece medidas procesales como la conservación rápida de datos almacenados, datos relativos al tráfico, la facultad de requerir la entrega de datos, el registro y aseguramiento de datos informáticos, la obtención de datos en tiempo real y la interceptación de datos relativos al contenido.

de conductas ilícitas cometidas mediante plataformas de IAc, especialmente en escenarios donde los proveedores y usuarios se encuentran en diferentes jurisdicciones.

Finalmente, cualquier estrategia de mejora debe incorporar un enfoque centrado en el menor (*child-centered approach*), que sitúe sus derechos, intereses y bienestar como eje principal del diseño, desarrollo y despliegue de la IAc. Esto supone reconocer que los menores son personas en proceso de formación cuyo entorno digital influye directamente en su desarrollo cognitivo, emocional y social. Un enfoque de este tipo obliga a repensar la tecnología como espacio de interacción que debe ser seguro, inclusivo y respetuoso con la dignidad y autonomía progresiva de la infancia y la adolescencia.

En síntesis, el cierre de los vacíos identificados y la mitigación de los riesgos detectados requiere una combinación equilibrada de medidas técnicas, jurídicas, educativas y de cooperación internacional. Ninguno de estos pilares puede actuar de forma aislada: la tecnología, por sí sola, no resolverá problemas de naturaleza social y ética; la legislación, sin apoyo técnico y educativo, será ineficaz; y la educación, sin un marco regulatorio y tecnológico adecuado, se encontrará en clara desventaja. Solo la articulación conjunta de estos elementos permitirá que la IAc se convierta en una herramienta que potencie las oportunidades de aprendizaje, creatividad y comunicación para los menores, sin poner en riesgo su integridad ni sus derechos fundamentales.

El examen de riesgos, vacíos regulatorios y necesidades de mejora desarrollado en la primera parte de esta monografía ha permitido establecer un marco de referencia sólido para comprender los desafíos que plantea la IAc en su interacción con niños y adolescentes.

Parte II

Esta segunda parte analiza las respuestas adoptadas por distintos ordenamientos jurídicos frente a los riesgos derivados de la interacción entre la inteligencia artificial conversacional y los menores. El estudio se centra en cuatro contextos normativos de referencia: Europa, Estados Unidos, China y Chile, que reflejan modelos regulatorios heterogéneos y permiten evaluar de manera crítica la adecuación de los marcos jurídicos vigentes para la protección de la infancia en entornos conversacionales mediados por inteligencia artificial.

II. Marco comparado de regulación de la “Inteligencia Artificial conversacional” y protección de menores

II.1. Unión Europea: marco normativo multinivel aplicable a la IAc y menores

El marco normativo de la Unión Europea en materia de protección de la infancia frente a la IAc se configura a partir de un entramado jurídico multinivel, en el que confluyen instrumentos vinculantes de derecho derivado, estrategias y recomendaciones no legislativas, y normas nacionales que desarrollan o concretan los estándares comunitarios. A diferencia del modelo estadounidense, basado en regulaciones sectoriales y competencias descentralizadas, y del modelo chino, desde un control estatal intensivo, la UE ha optado por una aproximación regulatoria integral que combina la protección de datos, la seguridad digital, la responsabilidad algorítmica y la protección de grupos vulnerables como ejes normativos fundamentales.⁶⁰

⁶⁰ El sistema europeo combina reglamentos directamente aplicables en todos los Estados miembros con directivas que requieren transposición nacional, y con estrategias no vinculantes que orientan las políticas públicas. Esta estructura busca garantizar que cualquier menor en la UE cuente con derechos equivalentes frente a riesgos digitales, aunque viva en países con distinta madurez tecnológica. El programa EU Kids Online ha mostrado que países con legislaciones más proactivas

El Reglamento (RGPD) constituye el pilar central de este marco. El art. 8 fija que, para el tratamiento de datos personales de menores en el contexto de servicios de la sociedad de la información basados en el consentimiento, el umbral de edad se sitúa en 16 años, aunque se permite a los Estados miembros reducirlo hasta un mínimo de 13. En consecuencia, el rango de protección oscila entre 13 y 16 años según el país.⁶¹ Este aspecto es especialmente relevante para la IAC, ya que condiciona la legalidad de la recopilación y tratamiento de datos conversacionales cuando el usuario es menor. Además, el RGPD impone principios de minimización de datos, limitación de la finalidad, exactitud e integridad, que obligan a los proveedores de Chatbots a implementar salvaguardas técnicas y organizativas adecuadas a los arts. 5 y 32. Cuando el tratamiento entrañe un alto riesgo para los derechos y libertades del menor, puede resultar preceptiva la realización de una evaluación de impacto relativa a la protección de datos, conforme al art. 35 del Reglamento.

En paralelo, la Directiva (UE) 2019/770 relativa a los contratos de suministro de contenidos y servicios digitales establece que los consumidores, incluidos los menores, tienen derecho a recibir información clara y precisa sobre el funcionamiento del servicio, las condiciones contractuales y los estándares de conformidad. Aunque no regula de

en línea con las estrategias de la Comisión Europea presentan menores niveles de exposición de niños a riesgos graves online, incluso sin leyes nacionales muy específicas para IAC.

⁶¹ El art. 8 del RGPD establece un rango flexible (13-16 años) para el consentimiento digital infantil, lo que crea una fragmentación práctica: un Chatbot que cumpla la ley en España (14 años) podría incumplirla en Países Bajos (16 años) si no adapta su verificación de edad.

manera directa los sistemas conversacionales, sí puede aplicarse a casos en los que un Chatbot se ofrece como parte de un servicio digital sujeto a contraprestación monetaria o en forma de datos personales.⁶²

La dimensión de seguridad digital infantil se refuerza con la Estrategia Europea para un Internet Mejor para los Niños (*Better Internet for Kids*, BIK), actualizada en 2022.⁶³ Este documento, no vinculante, propone líneas de acción como el diseño de entornos digitales seguros, la promoción de la alfabetización mediática, la capacitación parental y la creación de mecanismos de denuncia accesibles para menores.⁶⁴ Aunque sus disposiciones carecen de obligatoriedad jurídica, la BIK influye en la formulación de políticas nacionales y en la financiación de programas que pueden incluir la supervisión de la IAc.

Desde febrero de 2024, la Ley de Servicios Digitales (*Digital Services Act*, DSA)⁶⁵, introduce obligaciones vinculantes para intermediarios y plataformas en línea que operan en el mercado interior. Entre sus previsiones relevantes para la IAc destacan la prohibición de publicidad basada en perfiles cuando el destinatario sea un menor y la obligación de evaluar y mitigar riesgos sistémicos relacionados

⁶² Directive (EU) 2019/770, OJ L 136, 22.5.2019, 1-27.

⁶³ Una Década Digital para la infancia y la juventud: la nueva estrategia europea en favor de una Internet más adecuada para los niños (BIK +) 11 de mayo de 2022.

⁶⁴ La estrategia BIK ha permitido financiar centros nacionales, como Safer Internet Centres, que ofrecen asistencia técnica para configurar entornos seguros, incluidos filtros para Chatbots usados en colegios.

⁶⁵ Reglamento de servicios digitales (DSA). A partir del 17 de febrero de 2024, las normas del Reglamento de Servicios Digitales se aplican a todas las plataformas.

con la protección de la infancia. Estas obligaciones pueden resultar aplicables a proveedores de Chatbots cuando encajen en las categorías de servicios reguladas por el Reglamento, en función de su modelo de negocio y volumen de usuarios.

La Unión Europea adopta un enfoque basado en niveles de riesgo y establece obligaciones específicas para los sistemas que interactúan con personas vulnerables, categoría que incluye a los menores.⁶⁶ La aplicación del mencionado Reglamento (*AI Act*) a nivel nacional presenta diferencias. Por ejemplo, Francia aprobó en 2023 la Ley para la Regulación y Protección del Acceso en Línea de los Menores, que exige sistemas de verificación de edad para acceder a determinados contenidos y servicios digitales, incluidos potencialmente los Chatbots con funcionalidades sensibles.⁶⁷ España, por su parte, con la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales, ha fijado la edad de consentimiento en 14 años y ha introducido derechos específicos para los menores en entornos digitales, como la obligación de educar en el uso seguro de Internet.⁶⁸ Este margen de configuración

⁶⁶ Reglamento (UE) 2024/1689, cit.

⁶⁷ La Ley francesa de 2023 establece que cualquier servicio digital con contenidos potencialmente nocivos debe verificar la edad de los usuarios. Aunque su foco es el acceso a pornografía, se extiende a cualquier plataforma con contenido sensible, incluido el generado por IA (Loi n° 2023-566 du 7 juillet 2023).

⁶⁸ La Ley Orgánica 3/2018 obliga a que los centros educativos incluyan formación sobre uso seguro de Internet, lo que abre la puerta a capacitar en interacciones seguras con IAc. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, BOE-A-2018-16673.

nacional no altera la aplicabilidad directa del RGPD ni del AI Act, pero introduce diferencias prácticas en el régimen de consentimiento.

El marco europeo se distingue por el carácter integral, que combina protección de datos, derechos de los consumidores, seguridad digital y regulación específica de la IA. Sin embargo, persisten retos en la armonización de la edad de consentimiento, en la supervisión técnica efectiva de los sistemas conversacionales adaptativos y en la coordinación entre el RGPD, la DSA, el *AI Act* y las legislaciones nacionales. La plena efectividad de este modelo dependerá de la capacidad de la UE para desarrollar guías técnicas específicas, garantizar la cooperación transfronteriza y establecer mecanismos de auditoría según la complejidad y la rapidez de evolución de la IAc.⁶⁹

II.2. Estados Unidos: enfoque en la protección de datos infantiles

El marco normativo estadounidense en materia de protección de la infancia frente a la IAc se caracteriza por un enfoque descentralizado y sectorial, en el que las competencias regulatorias se reparten entre el Gobierno federal, los Estados y agencias administrativas con autoridad en ámbitos específicos. Aunque Estados Unidos ha desarro-

⁶⁹ El éxito del modelo europeo dependerá de integrar RGPD, DSA y Reglamento (UE) 2024/1689- [Reglamento de Inteligencia Artificial], cit., en procedimientos claros, con guías técnicas para auditar IAc en tiempo real: Veale et al. (2021, pp. 97-112).

llado instrumentos relevantes en la protección de datos infantiles en entornos digitales, no existe una legislación federal integral que regule de forma específica la inteligencia artificial conversacional ni sus riesgos asociados para menores.⁷⁰

El texto central en esta materia es la COPPA,⁷¹ promulgada en 1998 y gestionada por la “Federal Trade Commission” (en adelante, FTC). La ley establece que las empresas no pueden recopilar datos personales de menores de 13 años sin obtener un consentimiento verificable por parte de los padres o tutores. Esta obligación se extiende a sitios web, aplicaciones y servicios en línea dirigidos específicamente a menores, así como a aquellos que, sin estar diseñados para este público, tengan conocimiento efectivo de que están recogiendo información de menores.

Sin embargo, la COPPA⁷² presenta limitaciones notables. En primer lugar, su alcance se restringe a la protección de datos personales, sin contemplar expresamente otros riesgos propios de los entornos conversacionales, como la manipulación emocional, el fomento de conductas adictivas o la exposición a contenidos nocivos durante la interacción en tiempo real.⁷³ En segundo lugar, la edad de protección (menores de 13 años) no se ajusta a estándares internacionales como los de la Convención sobre los Derechos del Niño, que no establece un umbral de edad fijo y considera “niño” a toda persona menor de 18 años.⁷⁴

⁷⁰ Turow (2011).

⁷¹ COPPA. 16 CFR Part 312, 15, cit.

⁷² COPPA. 16 CFR Part 312, cit.

⁷³ Livingstone y Helsper (2007, pp. 671-696), cit.

⁷⁴ Asamblea General de las Naciones Unidas, Convención sobre los

Esto deja sin cobertura legal a adolescentes que, si bien poseen mayores competencias digitales, continúan siendo vulnerables a determinadas estrategias persuasivas utilizadas por sistemas de IA.

En años recientes, la FTC ha emitido orientaciones interpretativas para aplicar la COPPA a tecnologías emergentes, incluidos asistentes virtuales y sistemas basados en reconocimiento de voz. Sin embargo, estas guías no modifican el alcance material de la ley ni introducen obligaciones adicionales con rango legislativo. Junto a COPPA, Estados Unidos cuenta con legislaciones estatales que pueden complementar o reforzar la protección. La “California Consumer Privacy Act” (CCPA), en vigor desde 2020 y modificada por la “California Privacy Rights Act” (CPRA), amplía la edad de protección de datos hasta los 16 años y reconoce derechos adicionales, como el de limitar el uso de datos personales para fines de perfilado.⁷⁵ Aunque esta norma no está concebida específicamente para la IAc, sus disposiciones podrían aplicarse a Chatbots que operen en California y accedan a información personal de menores.

La coexistencia de normas federales y estatales genera un panorama regulatorio fragmentado, en el que el nivel de protección varía según la jurisdicción y el tipo de servicio ofrecido. A diferencia del modelo europeo, no existe una arquitectura normativa unificada que articule protección de datos, seguridad digital y obligaciones específicas relativas al diseño y funcionamiento de sistemas de inteligencia artificial.

Derechos del Niño, Resolución 44/25, 20 de noviembre de 1989, art. 1.

⁷⁵ “California Consumer Privacy Act”, Cal. Civ. Code 1798.100 (2018) y “California Privacy Rights Act”, Proposition 24 (2020).

En conjunto, el marco estadounidense ofrece un núcleo sólido en materia de consentimiento parental y limitaciones en la recogida de datos de menores de 13 años. No obstante, carece de un régimen integral que aborde de forma sistemática los riesgos propios de la inteligencia artificial conversacional, tales como la supervisión del contenido generado, la gestión de riesgos psicológicos o la imposición de estándares técnicos de diseño seguro. Esta ausencia normativa permite que Chatbots con capacidades adaptativas interactúen con menores más allá del ámbito estrictamente vinculado al tratamiento de datos personales.

II.3. China: modelo de control estatal y responsabilidad proactiva de plataformas

China presenta un modelo regulatorio de la IAc vinculado a una estrategia de gobernanza digital centralizada y a un marco normativo que combina control estatal de la información con políticas específicas de protección de menores. A diferencia de Estados Unidos, donde predomina la fragmentación normativa, el sistema chino se caracteriza por un alto grado de intervención gubernamental en la arquitectura tecnológica y en la imposición de obligaciones a proveedores y plataformas.

Uno de los pilares de este marco es la Ley de Protección de la Información Personal (“Personal Information Protection Law”, PIPL), en vigor desde noviembre de 2021. Inspirada en parte en el modelo europeo del RGPD, esta ley regula el tratamiento de datos personales y establece requisitos reforzados para la información de menores, con un umbral específico en el caso de menores de 14 años,

incluyendo el consentimiento informado de padres o tutores.⁷⁶ La PIPL introduce también la obligación de que las empresas minimicen la recogida de datos y adopten medidas de seguridad específicas. Esto afecta directamente a los Chatbots que procesan interacciones con usuarios infantiles.

A esta norma se suma la Ley de Protección de Menores (revisada en 2020), que dedica un bloque específico a la protección en entornos en línea. Entre sus disposiciones se incluyen obligaciones de prevención y control para proveedores de servicios digitales, con medidas orientadas a limitar la exposición a contenidos perjudiciales y a reducir riesgos asociados a hábitos de uso. El texto legal exige mecanismos de supervisión y respuesta ante contenidos o servicios que vulneren la normativa, con capacidad administrativa para ordenar retiradas o correcciones. Estas obligaciones pueden afectar a servicios conversacionales cuando operan como canales de acceso a contenidos o como interfaces de interacción.⁷⁷

En el ámbito específico de la IAc, China ha adoptado normas administrativas de alcance directo sobre servicios de IA generativa. La Regulación Provisional para Servicios de Inteligencia Artificial Generativa, emitida en julio de 2023 por la Administración del Ciberespacio de China (“Cyberspace Administration of China”, CAC), establece deberes de diligencia para los proveedores de servicios de

⁷⁶ “Personal Information Protection Law of the People’s Republic of China”, adoptada el 20 de agosto de 2021 en la trigésima reunión del Comité Permanente “National People’s Congress”.

⁷⁷ “Law of the People’s Republic of China on the Protection of Minors” (2020 Revision).

IA generativa, incluidos los Chatbots, con obligaciones relativas a la gestión de riesgos, a la seguridad del contenido y al cumplimiento de exigencias de gobernanza. El régimen impone controles de contenido alineados con criterios de “seguridad nacional” y “orden social”⁷⁸, y prevé obligaciones reforzadas cuando el destinatario es menor, con medidas técnicas para restringir acceso y reducir exposición a información inapropiada.

La combinación de estas normas configura un marco que integra supervisión estatal del contenido y responsabilidad proactiva de las plataformas. Este modelo puede reducir determinados riesgos evidentes, como pornografía, violencia, fraude o incitación a conductas delictivas, pero plantea tensiones respecto al acceso a información, al pluralismo y a la autonomía digital de los menores, en función de criterios estatales de control del contenido.

En términos de ejecución, China dispone de un sistema de inspección y sanción con capacidad administrativa intensa. Las plataformas que incumplen pueden enfrentarse a multas, suspensión de servicios o inhabilitación operativa. La CAC y otras autoridades sectoriales cuentan con facultades para exigir medidas correctivas y revisar el cumplimiento de obligaciones técnicas, con un nivel de intervención más elevado que el habitual en modelos occidentales. En este contexto, los Chatbots deben ajustarse a requisitos de conformidad que alcanzan al tratamiento de datos, al filtrado de contenidos y a los mecanismos de control de acceso.

En síntesis, el marco normativo chino aplicable a la

⁷⁸ “Cyberspace Administration of China, Interim Administrative Measures for Generative Artificial Intelligence Services” (2023).

interacción entre IAc y menores combina protección de datos personales, obligaciones de seguridad y controles de contenido bajo supervisión administrativa fuerte. Su eficacia para reducir riesgos visibles puede resultar alta, pero el modelo introduce interrogantes sobre el equilibrio entre protección, derechos digitales y libertad de información, además de límites técnicos persistentes en verificación de edad y prevención de suplantación de identidad.

II.4. Chile: marco emergente y vacíos regulatorios en materia de IAc y menores

El caso chileno constituye un marco regulatorio emergente, todavía en proceso de consolidación, que integra instrumentos generales de protección de derechos e iniciativas orientadas a la gobernanza de la IA. En contraste con los marcos de la Unión Europea o chinos, Chile carece, por el momento, de una legislación integral aplicable a la IAc, lo que sitúa la protección de los menores en un contexto normativo parcialmente incompleto y sujeto a cambios.

En la actualidad, la protección de la infancia frente a riesgos digitales se articula principalmente a través de normas de alcance general en materia de datos personales, derechos de la niñez, consumo y delitos informáticos.⁷⁹ No obstante, estas disposiciones generales no abordan de forma específica los riesgos derivados de la interacción conversacional con sistemas de IA, como la manipulación cognitiva, la personalización persuasiva o la generación

⁷⁹ Ley chilena n.º 19.628, sobre protección de la vida privada.

automatizada de contenidos. En este sentido, la aplicabilidad a Chatbots suele depender de interpretaciones por analogía, sin obligaciones técnicas específicas de diseño o auditoría.

En el ámbito penal, la Ley n.º 21.459 sobre Delitos Informáticos (2022), que derogó la Ley n.º 19.223 y modificó diversos cuerpos legales con el fin de adecuar el ordenamiento chileno al Convenio de Budapest sobre Ciberdelincuencia, constituye un instrumento relevante. Esta norma tipifica conductas como el acceso ilícito, la interceptación indebida, el ataque a la integridad de sistemas informáticos y el uso fraudulento de datos. Si bien fortalece la respuesta penal frente a conductas delictivas cometidas a través de medios digitales, su enfoque es reactivo y no incorpora obligaciones preventivas relativas al diseño o funcionamiento de sistemas de IAc.

El marco normativo chileno presenta limitaciones estructurales para abordar las especificidades de la IAc. Esta insuficiencia se vincula tanto a la antigüedad de la Ley n.º 19.628 sobre protección de la vida privada (Chile, 1999) como a su escasa adecuación a los estándares internacionales contemporáneos en materia de gobernanza algorítmica y protección reforzada de menores en entornos digitales. A diferencia del Reglamento General de Protección de Datos europeo, que incorpora evaluaciones de impacto y obligaciones específicas para tratamientos de alto riesgo, la legislación chilena no contempla categorías normativas relativas al tratamiento automatizado avanzado ni establece obligaciones diferenciadas cuando el titular de los datos es menor de edad. Tampoco regula de manera expresa el perfilado algorítmico ni la toma de decisiones automatizadas en contextos conversacionales.

En el ámbito de la protección frente a riesgos digitales graves, el ordenamiento chileno ha reforzado su marco penal mediante la Ley n.º 21.459 sobre Delitos Informáticos (Chile, 2022), que adecuó la legislación interna al Convenio sobre la Ciberdelincuencia (Consejo de Europa, 2001). Esta norma tipifica conductas como el acceso ilícito, la interceptación indebida de datos y el ataque a la integridad de sistemas informáticos. Asimismo, el Código Penal chileno contempla figuras vinculadas al *grooming* y a la explotación sexual infantil. Sin embargo, este conjunto normativo responde a una lógica de reacción *ex post* frente al daño consumado. No establece estándares preventivos de diseño seguro ni obligaciones técnicas dirigidas a desarrolladores y proveedores de sistemas de IAc.

Desde una perspectiva comparada, la ausencia de mecanismos obligatorios de evaluación de riesgos, auditoría algorítmica o verificación robusta de edad sitúa a Chile en un escenario de regulación fragmentaria. Existen disposiciones aplicables de forma indirecta en materia de datos personales, ciberdelincuencia o protección al consumidor, pero no se configura un régimen sistemático orientado a la interacción entre menores y sistemas conversacionales basados en aprendizaje automático.

El contraste comparado refuerza esta conclusión. Mientras que Estados Unidos dispone de un marco federal específico centrado en la privacidad infantil, la “Children’s Online Privacy Protection Act” (15 U.S.C. 6501–6506), aunque limitado a menores de trece años, y China ha desarrollado un modelo de supervisión administrativa intensa a través de la “Personal Information Protection Law” (“Standing Committee of the National People’s Congress”, 2021) y regulaciones específicas sobre IA generativa (“Cy-

berspace Administration of China”, 2023), Chile permanece en una fase de formulación normativa incipiente. Esta situación plantea un desafío relevante en términos de protección efectiva, pero también ofrece la posibilidad de diseñar un marco que incorpore estándares internacionales emergentes en materia de evaluación de impacto, transparencia algorítmica y salvaguardas reforzadas para menores (UNICEF, 2020).

En ausencia de obligaciones explícitas relativas a auditorías de contenido, supervisión humana obligatoria o mecanismos avanzados de verificación de edad, el modelo chileno descansa en un amplio margen de autorregulación. La evidencia comparada sugiere que este enfoque presenta limitaciones cuando se trata de sistemas conversacionales que evolucionan mediante aprendizaje automático, donde los riesgos pueden manifestarse de manera progresiva y afectar de forma desproporcionada a usuarios en situación de vulnerabilidad (OECD, 2021).⁸⁰

⁸⁰ “United Nations Children’s Fund” – UNICEF (2021). Policy Guidance on AI for Children (Version 2.0). Recommendations for building AI policies and systems that uphold child rights.

II.5. Necesidades y oportunidades de mejora frente a los vacíos regulatorios de la Inteligencia Artificial conversacional

La presencia de la IAc en los entornos cotidianos de niños y adolescentes plantea desafíos que los marcos jurídicos y sociales actuales no abordan de forma suficiente. Los vacíos regulatorios afectan a dimensiones técnicas, educativas y comunitarias que condicionan la protección efectiva de los menores.

Entre estas carencias se incluyen la falta de estándares de seguridad aplicables a sistemas conversacionales accesibles a menores, la debilidad de los mecanismos de verificación de edad, la ausencia de auditorías independientes y la escasa integración de competencias digitales críticas en los procesos educativos. A ello se suma una desigual capacidad de mediación por parte de las familias, que incrementa la exposición a riesgos como la manipulación cognitiva y emocional, la explotación de datos o la generación de vínculos de dependencia con sistemas que simulan empatía.⁸¹

Las consideraciones siguientes son un conjunto de necesidades y oportunidades de mejora adaptables a distintos contextos normativos y culturales. El objetivo es doble: presentar una propuesta que pueda reforzar la protección de la infancia y, al mismo tiempo, promover un uso seguro, crítico y provechoso de la IAc, evitando respuestas exclusivamente prohibicionistas.

Desde esta perspectiva, se identifican cinco líneas estratégicas interconectadas: el ámbito legislativo y regulatorio,

⁸¹ Arnd-Caddigan (2015), cit.

el ámbito educativo, el ámbito social y comunitario, el ámbito familiar y los mecanismos de supervisión y auditoría. Su eficacia depende de una articulación coherente que evite soluciones aisladas y permita responder a la complejidad de los entornos conversacionales mediados por IA.

Antes de abordar cada ámbito de intervención, conviene subrayar que estas oportunidades de mejora no operan de forma aislada. El refuerzo del marco normativo pierde eficacia si no se acompaña de un esfuerzo educativo sostenido, del mismo modo que la mediación familiar y la acción comunitaria requieren respaldo institucional y jurídico para resultar efectivas.

II.5.1. Ámbito legislativo y regulatorio general

Este apartado analiza el marco legislativo general aplicable a la IAc en contextos de infancia desde una perspectiva comparada. Se examinan distintos modelos regulatorios con el fin de identificar convergencias, divergencias y vacíos estructurales que afectan a la protección de niños y adolescentes.

El primer nivel de protección frente a los riesgos de la IAc en la infancia es la ley. Sin embargo, la eficacia de la legislación depende de que esta sea clara, actualizada y técnicamente ajustada al objeto que regula.

La experiencia comparada muestra un espectro muy amplio: desde marcos exhaustivos como el de China, que combinan control estatal y exigencias técnicas estrictas, hasta sistemas más fragmentarios como el de Estados Unidos, centrados casi en exclusiva en la protección de datos

a través de la COPPA.⁸² En América Latina, el panorama es todavía más heterogéneo, con países que han formulado estrategias nacionales de IA sin traducción normativa concreta, y otros que mantienen legislaciones obsoletas en materia de privacidad.

La oportunidad de mejora en este ámbito pasa, en primer lugar, por integrar disposiciones específicas sobre Chatbots y menores en las leyes actuales: protección de datos, delitos informáticos, derechos del consumidor, y, sobre todo, en leyes de protección de la infancia.

Pero la Ley no puede limitarse a regular el uso de datos o la prohibición de contenidos ilícitos. Debe incluir estándares técnicos vinculantes para el diseño seguro de Chatbots accesibles a menores. Esto implica definir protocolos de filtrado de contenido, auditorías de sesgos, verificación rigurosa de edad y limitaciones en la personalización basada en datos sensibles. En esta línea, organismos internacionales como UNICEF (2020) han propuesto principios específicos para la IA dirigida o accesible a menores, que incluyen la evaluación de impacto sobre los derechos del niño, la transparencia adaptada⁸³ a la edad y la incorporación del interés superior del menor como criterio de diseño.

La ausencia de estos estándares es un vacío crítico que permite la circulación de productos con riesgos potenciales sin pasar por un control previo similar al que se exige, por ejemplo, en juguetes físicos.

La cooperación internacional es otro pilar imprescindi-

⁸² COPPA. 16 CFR Part 312, cit.

⁸³ “United Nations Children’s Fund” – UNICEF (2020). Policy Guidance on AI for Children, cit.

ble. La naturaleza transfronteriza de los Chatbots permite que un vacío en una jurisdicción pueda convertirse en una puerta de entrada de posibles riesgos. En este sentido, se podrían aprovechar foros como la OECD o el Consejo de Europa para promover protocolos comunes y reconocimiento mutuo de auditorías, del mismo modo que sucede con el Convenio de Budapest⁸⁴ para delitos informáticos.

Los vacíos regulatorios identificados en el análisis comparado pueden interpretarse también como oportunidades para una mejora estructural de los marcos de protección existentes. La ausencia de una regulación específica sobre IAc en contextos de infancia sugiere la conveniencia de incorporar disposiciones expresas relativas a Chatbots en las normas sobre protección de menores, protección de datos y delitos informáticos, bajo el liderazgo de los legisladores y de los organismos especializados en infancia.

De igual modo, la falta de estándares técnicos vinculantes pone de relieve la necesidad de desarrollar normas de alcance internacional, adaptadas a sistemas conversacionales de uso educativo y recreativo, en cuya elaboración deberían participar tanto las autoridades competentes en materia tecnológica como los organismos de normalización. En el plano internacional, la carencia de mecanismos específicos de cooperación frente a delitos facilitados por inteligencia artificial apunta a la posibilidad de ampliar convenios ya existentes, incorporando de forma explícita

⁸⁴ Consejo de Europa, Convenio sobre la Ciberdelincuencia (Convenio de Budapest), 2001. Véase Instrumento de Ratificación del Convenio sobre la Ciberdelincuencia (Budapest, 23 de noviembre de 2001, publicado en España, BOE n.º 226, de 17 de septiembre de 2010, pp.78847-78896).

estas nuevas modalidades de riesgo.

Finalmente, las debilidades actuales en los sistemas de verificación de edad evidencian la urgencia de implantar mecanismos auditables que integren salvaguardas de privacidad y protección de datos, con una responsabilidad compartida entre proveedores de servicios y agencias reguladoras. La efectividad de estas medidas depende, en todo caso, de una coordinación real entre actores públicos y privados, capaz de traducir los principios normativos en prácticas operativas.

II.5.2. Ámbito educativo

El sistema educativo constituye uno de los espacios con mayor capacidad de influencia sobre la forma en que niños y adolescentes interpretan, evalúan y normalizan la interacción con la IAc. A diferencia del marco normativo, que establece límites externos, la educación actúa sobre procesos internos de comprensión, juicio y autorregulación. Por este motivo, los entornos educativos formales e informales desempeñan un papel estructural en la prevención de riesgos asociados al uso de Chatbots.⁸⁵

Uno de los principales déficits detectados en muchos sistemas educativos es que la alfabetización digital continúa centrada en competencias instrumentales básicas. Saber buscar información o manejar herramientas digitales resulta necesario, pero insuficiente ante sistemas conversacionales capaces de personalizar el discurso, recordar in-

⁸⁵ Livingstone and Helsper (2007, pp. 671-696), cit.

teracciones previas y modular sus respuestas en función del interlocutor. Desde una perspectiva psicológica, estas características favorecen la atribución de competencia, neutralidad y autoridad al Chatbot.

Diversos estudios han mostrado que niños y adolescentes tienden a confiar en sistemas conversacionales incluso cuando no comprenden su funcionamiento interno. Esto refuerza la necesidad de una alfabetización digital crítica que cuestione esta autoridad percibida y que permita comprender cómo funcionan estos sistemas, identificar sesgos, cuestionar la veracidad de las respuestas y mantener una distancia reflexiva frente a su apariencia de interlocutor humano.⁸⁶

Las experiencias internacionales muestran que este enfoque es viable cuando se integra de forma transversal. Iniciativas educativas desarrolladas en países como Finlandia han incorporado contenidos sobre funcionamiento básico de la IA, con el programa *Elements of AI*,⁸⁷ originalmente diseñado para adultos, adaptado para adolescentes, con módulos específicos sobre ética, sesgos y funcionamiento básico de los sistemas de IA.

El profesorado ocupa una posición clave en este proceso. Sin embargo, numerosos estudios señalan que muchos docentes carecen de formación específica en inteligencia

⁸⁶ “Department of Homeland Security” (DHS), Stop.Think. Connect Campaign, National cybersecurity awareness Campaign (2010).

⁸⁷ Es un programa desarrollado por la University of Helsinki en colaboración con el gobierno de Finlandia. Su objetivo es difundir los conocimientos fundamentales de la IA a toda la población, independientemente de su formación técnica, de modo que los participantes puedan comprender el funcionamiento de los sistemas de IA, sus posibles aplicaciones y las implicaciones éticas y sociales asociadas.

artificial y en sus implicaciones para la infancia. Esta carencia limita su capacidad para acompañar críticamente el uso educativo de Chatbots. La formación docente resulta eficaz cuando no se reduce al aprendizaje técnico de la herramienta, sino que incorpora una comprensión de los riesgos psicológicos, sociales y éticos, junto con estrategias de mediación en el aula. De este modo, el docente puede guiar al alumnado en la interpretación de las respuestas generadas por la IA, fomentar el contraste con otras fuentes y convertir la interacción en un objeto de análisis compartido.

Desde el punto de vista del desarrollo del menor, las competencias educativas relevantes pueden entenderse como un conjunto integrado. Incluyen la capacidad de evaluar críticamente la información recibida, reconocer los límites de la inteligencia artificial, respetar la privacidad propia y ajena y actuar de forma responsable ante interacciones problemáticas. Estas dimensiones cognitivas, éticas y operativas deben integrarse en actividades que reproduzcan situaciones reales de uso, favoreciendo una comprensión situada y significativa.

El vínculo entre escuela y familia refuerza este aprendizaje. Cuando los mensajes educativos se comparten y se explicitan fuera del aula, aumenta la coherencia normativa y se reduce la disonancia entre entornos. Talleres conjuntos, actividades participativas y proyectos presentados por el propio alumnado facilitan que las competencias desarrolladas en la escuela se consoliden en el hogar.

En conjunto, el ámbito educativo configura la relación a largo plazo de los menores con la tecnología. Una educación orientada a la alfabetización crítica reduce la vulnerabilidad frente a la manipulación y la dependencia emocional de sistemas conversacionales, y contribuye a formar

ciudadanos capaces de cuestionar, exigir transparencia y participar de forma consciente en entornos digitales mediados por IA.

La adquisición de estas competencias resulta especialmente eficaz cuando se apoya en estrategias educativas activas que sitúan al alumnado ante situaciones reales de interacción con sistemas conversacionales. El análisis crítico de respuestas generadas por Chatbots, el contraste con fuentes fiables y la reflexión compartida sobre los resultados permiten que el uso de la inteligencia artificial se convierta en un objeto de aprendizaje y no en una autoridad incuestionada.

Desde esta perspectiva, el ámbito educativo contribuye a construir un marco de referencia duradero para la vida digital del menor. Una alfabetización crítica sólida reduce la vulnerabilidad frente a la manipulación, el perfilado psicológico no consentido y la dependencia emocional de sistemas conversacionales, y prepara a niños y adolescentes para exigir transparencia, responsabilidad y seguridad a los actores que desarrollan estas tecnologías.

II.5.3. Ámbito social y comunitario

Más allá de la escuela y del marco normativo, la protección de niños y adolescentes frente a los riesgos de la IAc se articula también en el espacio social y comunitario. Este ámbito comprende el conjunto de actores y estructuras que influyen en la vida cotidiana de los menores sin formar parte directa del sistema educativo o legislativo. Asociaciones familiares, organizaciones sociales, instituciones culturales, medios de comunicación y autoridades locales

contribuyen de forma decisiva a definir qué usos de la tecnología se consideran aceptables, problemáticos o deseables dentro de una comunidad.⁸⁸

Desde una perspectiva sociológica, este entorno cumple una función clave de “normalización”. Las prácticas digitales se interiorizan mediante la observación de comportamientos, discursos y respuestas colectivas ante los riesgos. El ámbito comunitario actúa, en este sentido, como un espacio intermedio entre la regulación institucional y la experiencia individual, con capacidad para detectar tempranamente cambios de conducta, dinámicas de uso problemáticas o nuevas formas de interacción con sistemas conversacionales.

Por ejemplo, campañas como *Stop, Think, Connect* en Estados Unidos han logrado combinar mensajes preventivos con recursos educativos prácticos, dirigidos tanto a menores como a adultos.⁸⁹

La prevención en este nivel se apoya en tres dimensiones interrelacionadas: concienciación, participación y vigilancia social. La concienciación implica construir un discurso público equilibrado que permita comprender tanto los riesgos como las condiciones de un uso seguro. Iniciativas de comunicación social, como las desarrolladas en campañas públicas en distintos países, muestran que la prevención resulta más eficaz cuando evita el alarmismo y ofrece criterios claros de interpretación y actuación.

La participación comunitaria refuerza este proceso al implicar a los propios miembros de la comunidad en la transmisión de normas y valores digitales. Las experiencias

⁸⁸ Livingstone and Blum-Ross (2020).

⁸⁹ Livingstone and Helsper (2008, pp. 581-599).

de educación entre iguales y las iniciativas impulsadas por asociaciones familiares o juveniles favorecen la identificación y la confianza, y permiten que los mensajes preventivos se adapten a contextos sociales y culturales concretos.

Iniciativas de *peer education*, donde jóvenes formados instruyen a sus iguales sobre el uso seguro de Chatbots, han mostrado resultados prometedores en entornos urbanos de Chile y Argentina, al fomentar la identificación entre formador y receptor.⁹⁰ De igual modo, las asociaciones de padres pueden convertirse en canales efectivos para distribuir guías de seguridad, organizar charlas con expertos en IA y coordinar acciones con autoridades locales.

La vigilancia social completa este entramado. A través de mecanismos formales e informales, la comunidad puede identificar incidentes, patrones de riesgo o prácticas emergentes relacionadas con la interacción de menores con Chatbots. Observatorios locales, redes de alerta temprana y espacios de coordinación permiten canalizar esta información hacia respuestas proporcionadas, evitando tanto la inacción como la excesiva reacción.⁹¹

En este sentido, la protección frente a los riesgos de la IAc deja de ser una suma de intervenciones aisladas y se configura como una responsabilidad compartida, sostenida por la interacción coherente entre familia, comunidad, instituciones y actores locales.

⁹⁰ Arnd-Caddigan (2015, pp. 247-248), cit.

⁹¹ “eSafety Commissioner” (2023).

II.5.4. Ámbito familiar

El hogar constituye el primer espacio en el que niños y adolescentes construyen su relación con la tecnología y, de forma cada vez más habitual, con sistemas de IAc. Antes de cualquier mediación escolar o comunitaria, muchos menores interactúan ya en casa con asistentes virtuales, *bots* integrados en aplicaciones de mensajería o personajes conversacionales presentes en videojuegos y plataformas de ocio digital.

Estas primeras interacciones no son neutras. Contribuyen a configurar la forma en que el menor percibe la IA como interlocutor, fuente de información o figura de referencia. Cuando el contacto se produce de manera temprana y sin acompañamiento adulto, el sistema conversacional puede ser interiorizado como una presencia cotidiana fiable, accesible y carente de conflicto, lo que influye en la construcción del criterio propio y en la atribución de autoridad a la respuesta algorítmica.

Desde una perspectiva psicológica, este entorno favorece dinámicas de familiaridad y confianza que no siempre se corresponden con una comprensión real del funcionamiento del sistema. El menor tiende a interpretar la coherencia lingüística y la disponibilidad constante del Chatbot como señales de competencia y neutralidad, lo que puede dificultar la distinción entre orientación técnica, opinión simulada y acompañamiento humano. Estas percepciones iniciales, adquiridas en el ámbito doméstico, condicionan de manera duradera la forma en que niños y adolescentes se relacionan con la inteligencia artificial en otros contex-

tos educativos y sociales.⁹²

El papel de la familia en la protección frente a los riesgos de la IAc no consiste en imponer prohibiciones ni en limitar el tiempo de uso de forma aislada. Su función principal es ejercer una mediación activa, entendida como un acompañamiento continuo en el que participan en la experiencia digital del menor, dialogan sobre las interacciones mantenidas con sistemas conversacionales y establecen criterios compartidos de uso.

Desde una perspectiva psicológica, esta mediación favorece el desarrollo del juicio crítico y de la autorregulación. El menor aprende a interpretar las respuestas del Chatbot como productos técnicos y no como verdades incuestionables, y a reconocer los límites de una interacción que simula comprensión sin implicar una relación humana real. La norma deja de percibirse como una imposición externa y se percibe como un marco de referencia comprensible y negociado.

Este modelo de acompañamiento refuerza la autonomía progresiva y reduce la atracción por el uso oculto o transgresor de la tecnología. Al sentirse escuchado y orientado, el menor tiende a compartir dudas, experiencias problemáticas o situaciones incómodas derivadas de la interacción con la inteligencia artificial, lo que permite una intervención temprana y una protección más eficaz en el entorno familiar.⁹³

La mediación activa en el entorno familiar se articula a través de prácticas complementarias que inciden directa-

⁹² UK Safer Internet Centre”, Resources for Parents and Carers (2022).

⁹³ Yeung (2018, pp. 505-523), cit.

mente en la forma en que el menor construye sentido, criterio y vínculo con la IAc. Compartir la experiencia digital implica acompañar al menor durante la interacción con el Chatbot, observar el tipo de preguntas que formula y las respuestas que recibe, y comentar de manera abierta lo que ocurre. Esta presencia adulta crea un espacio de reflexión inmediata que favorece el desarrollo del pensamiento crítico y la evaluación ética de la información generada.

Explicar las capacidades y los límites de la inteligencia artificial cumple una función cognitiva esencial. Al comprender que el sistema produce respuestas a partir de patrones aprendidos, el menor aprende a relativizar la autoridad del discurso algorítmico. Esta comprensión reduce la tendencia a atribuir intencionalidad, juicio moral o fiabilidad absoluta a un interlocutor que presenta una apariencia lingüística convincente, pero carece de conciencia y de responsabilidad.

La negociación de reglas de uso introduce una dimensión relacional clave. Acordar criterios sobre tiempos, temáticas y formas de interacción permite que las normas se interioricen como marcos compartidos y no como imposiciones externas. Este proceso fortalece la autorregulación y facilita que el menor comunique situaciones de incomodidad o duda sin temor a sanciones automáticas.

Junto a estas prácticas, la alfabetización emocional adquiere un papel central. Muchos sistemas conversacionales están diseñados para simular empatía, recordar intercambios previos y adaptar su lenguaje al estado emocional del usuario. Estas dinámicas pueden resultar funcionales en contextos educativos o de apoyo supervisado, pero también generan riesgos específicos. En ausencia de mediación, el menor puede desarrollar vínculos de dependencia

o confusión respecto a la naturaleza no humana del interlocutor, especialmente cuando la interacción responde a necesidades afectivas. Reconocer esta dimensión emocional resulta decisivo para proteger el desarrollo psicológico y relacional en entornos mediados por IA.

Uno de los principales obstáculos para la mediación familiar es la brecha de competencias digitales entre generaciones. En hogares donde los adultos carecen de una alfabetización tecnológica suficiente, la supervisión pierde eficacia y el menor tiende a desenvolverse con mayor autonomía técnica que criterio crítico. Esta asimetría altera la relación educativa, ya que dificulta que el adulto actúe como referente y favorece la delegación implícita del juicio en el sistema conversacional.

Desde una perspectiva sociológica, esta situación debilita la función protectora del entorno familiar y aumenta la dependencia del menor respecto a fuentes externas de orientación, incluidas las ofrecidas por la propia inteligencia artificial. La falta de competencias reduce la confianza del adulto para intervenir, dialogar o cuestionar el uso que el menor hace de estas tecnologías.

Las iniciativas formativas dirigidas a padres y cuidadores, desarrolladas en países como Australia o España, responden a esta carencia desde un enfoque preventivo. Estos programas, ofrecidos con frecuencia en espacios comunitarios, refuerzan la capacidad de los adultos para comprender el funcionamiento de los Chatbots, interpretar sus condiciones de uso y acompañar al menor en la protección de su privacidad. Más allá del aprendizaje técnico, estas formaciones contribuyen a restablecer la mediación familiar como espacio de diálogo, confianza y orientación crítica en el uso cotidiano de la IA.

La coherencia entre familia, escuela y comunidad resulta decisiva para que las normas y valores transmitidos en el hogar se consoliden en otros entornos de socialización. Cuando niños y adolescentes reciben mensajes consistentes sobre el uso de la IAc, aumenta la probabilidad de que los interioricen y los incorporen a su conducta cotidiana. Esta continuidad reduce la confusión normativa y refuerza la construcción de límites comprensibles y estables.

El ámbito familiar desempeña también un papel central en la detección temprana de situaciones problemáticas. Al tratarse del espacio de mayor convivencia y confianza, es en el hogar donde suelen manifestarse primero cambios de conducta que pueden estar vinculados a una interacción negativa con sistemas conversacionales. El aislamiento progresivo, la ansiedad, las alteraciones del sueño o las variaciones bruscas en el uso de dispositivos actúan como señales de alerta que requieren atención y acompañamiento.

Desde una perspectiva preventiva, la capacidad de la familia para reconocer estas señales y activar respuestas resulta determinante. Más allá de medidas técnicas puntuales, lo relevante es disponer de un marco de actuación compartido, que permita dialogar sobre lo ocurrido, ajustar el uso de la tecnología y, cuando sea necesario, recurrir a apoyo profesional o institucional. Esta respuesta temprana reduce la escalada del problema y protege el bienestar emocional del menor.

Una familia que se implica se forma y coordina con la escuela y la comunidad se convierte en un espacio de contención y orientación estable. En este contexto, la IAc puede integrarse en la vida del menor como una herramienta más, sin desplazar las relaciones humanas ni erosionar los procesos de socialización y desarrollo personal.

II.5.5. Mecanismos de supervisión y auditoría

Todas las estrategias de protección frente a los riesgos de la IAc en la infancia comparten una limitación estructural: dependen de que existan mecanismos efectivos que garanticen su cumplimiento. Sin supervisión ni auditoría, las normas y las recomendaciones pierden capacidad transformadora y se convierten en marcos declarativos sin impacto real en la experiencia cotidiana de niños y adolescentes. Por este motivo, este ámbito atraviesa todos los anteriores y constituye uno de los más complejos de articular en entornos digitales de alcance global.

Desde una perspectiva funcional, la supervisión y la auditoría responden a lógicas complementarias. La supervisión permite un seguimiento continuado del comportamiento de los sistemas conversacionales y facilita la detección temprana de interacciones problemáticas. La auditoría introduce una evaluación periódica y estructurada, orientada a analizar si el diseño, el funcionamiento y el uso de estos sistemas respetan criterios técnicos, éticos y de protección de derechos. En el caso de los Chatbots accesibles a menores, esta combinación resulta especialmente relevante, ya que los riesgos no siempre se manifiestan de forma inmediata ni visible.

Los mecanismos de control basados exclusivamente en la autorregulación presentan límites evidentes. Cuando la supervisión depende únicamente del proveedor tecnológico, la lógica comercial puede interferir en la identificación de fallos o en la adopción de medidas correctivas. Desde una perspectiva sociológica, esta situación erosiona la confianza pública y refuerza la percepción de opacidad, especialmente cuando los sistemas influyen en procesos sensi-

bles como la formación de criterio, la gestión emocional o la socialización digital del menor.

Las respuestas institucionales a este desafío varían según los contextos normativos y culturales. Algunos sistemas priorizan mecanismos formales de evaluación y control, mientras que otros confían en mayor medida en la responsabilidad empresarial. Estas diferencias reflejan concepciones distintas sobre el papel del Estado y sobre el grado de intervención legítima en la protección de la infancia frente a tecnologías complejas. En ausencia de controles verificables, el riesgo tiende a desplazarse hacia los usuarios más vulnerables.

Desde un punto de vista social, la supervisión eficaz adopta una estructura multinivel. A la responsabilidad de los desarrolladores y a la intervención de las autoridades se añaden las auditorías independientes y la capacidad de detección temprana de organizaciones sociales y observatorios ciudadanos. Este entramado resulta funcional cuando existe un flujo de información transparente y accesible, que permita transformar señales dispersas en respuestas coordinadas y comprensibles para familias y educadores.

En el plano normativo, la Unión Europea, a través del RGPD y de la Ley de la IA,⁹⁴ ha introducido la obligación de realizar evaluaciones de impacto y auditorías para los sistemas clasificados como de alto riesgo, entre los que podrían incluirse los Chatbots con acceso directo a menores.⁹⁵ China, por su parte, ha establecido un marco que permite a las autoridades acceder al código y a los datos de en-

⁹⁴ Reglamento (UE) 2024/1689 [Reglamento de Inteligencia Artificial], cit.

⁹⁵ Yeung (2018, pp. 505-523), cit.

trenamiento de los sistemas para verificar su conformidad con la ley, retirando inmediatamente del mercado aquellos que incumplan la exigencias de seguridad o contenido. En Estados Unidos, la COPPA establece controles sobre el uso de datos de menores, pero no impone un sistema generalizado de auditorías técnicas, lo que deja un margen amplio de autorregulación a las empresas.⁹⁶

La dificultad de supervisar y auditar sistemas de IAC a escala global deriva de su carácter transfronterizo. Un mismo servicio puede operar simultáneamente en contextos jurídicos muy distintos, lo que limita la eficacia de los controles exclusivamente nacionales. Para afrontar este problema, una vía prometedora es la cooperación internacional mediante acuerdos de reconocimiento mutuo de auditorías, basados en estándares comunes y en la acreditación recíproca de entidades evaluadoras.

Desde esta perspectiva, la supervisión de la IAC en contextos de infancia puede articularse como un sistema multinivel. En un primer plano, los propios desarrolladores y proveedores deben asumir responsabilidades de control interno, mediante la monitorización de interacciones y el registro de incidentes relevantes. A este nivel se suma la supervisión institucional ejercida por las autoridades competentes en materia de protección de datos, regulación tecnológica y telecomunicaciones, con capacidad para auditar el cumplimiento normativo y aplicar sanciones. De forma complementaria, las auditorías independientes realizadas por entidades acreditadas permiten evaluar el cumplimiento de estándares técnicos y éticos con criterios de impar-

⁹⁶ Livingstone y Helsper (2007, pp. 671-696).

cialidad, mientras que los observatorios ciudadanos y las organizaciones sociales aportan una detección temprana de riesgos desde el terreno.

Este modelo solo resulta eficaz si existe una infraestructura sólida de intercambio de información. Los incidentes detectados en el ámbito comunitario deben poder comunicarse con rapidez a las autoridades competentes y, cuando proceda, activar mecanismos de auditoría independiente. Del mismo modo, los resultados de las evaluaciones periódicas deberían hacerse públicos, al menos de forma parcial, para facilitar que familias, educadores y comunidades conozcan el grado de seguridad de los sistemas utilizados por los menores.

Una línea de mejora especialmente relevante consiste en integrar la perspectiva de los propios niños y adolescentes en los procesos de supervisión. Los menores pueden identificar problemas de uso que no siempre resultan visibles para los adultos y formular propuestas concretas que refuerzan la seguridad sin impedir un uso legítimo de la tecnología.

En definitiva, los mecanismos de supervisión y auditoría constituyen un pilar central para que las políticas de protección infantil en el ámbito de la IAc se traduzcan en prácticas efectivas. Su eficacia depende de la combinación de independencia, capacidad técnica, coordinación internacional y apertura a la participación ciudadana, como condiciones necesarias para pasar de los principios normativos a una protección real y verificable.

II.5.6. Síntesis e integración de las oportunidades de mejora

Las oportunidades de mejora identificadas deben entenderse como componentes de una arquitectura integral de protección frente a los vacíos regulatorios y los ciberpe- ligros vinculados a la inteligencia artificial conversacional en la infancia y la adolescencia. Su eficacia no depende de la fortaleza aislada de cada ámbito de intervención, sino de su articulación coherente como un sistema de garantías interconectadas.

En el plano legislativo y regulatorio, la carencia de dis- posiciones específicas y de estándares técnicos vinculantes aplicables a Chatbots accesibles a menores revela una debi- lidad estructural. La existencia de normas generales resulta insuficiente cuando no se acompaña de capacidad institu- cional de supervisión, mecanismos de auditoría indepen- dientes y actualización periódica conforme a la evolución tecnológica. La efectividad del marco jurídico depende, por tanto, de su traducción operativa en controles verificables.

El ámbito educativo introduce la dimensión preventiva de mayor alcance. La alfabetización digital crítica fortalece la autonomía intelectual del menor y reduce su vulnerabili- dad frente a la manipulación, la desinformación o la depen- dencia tecnológica. Esta protección se consolida cuando escuela, familia y entorno comunitario transmiten criterios coherentes sobre el uso de sistemas conversacionales.

El ámbito social y comunitario cumple una función de enlace y refuerzo normativo. Su proximidad a la expe- riencia cotidiana permite detectar dinámicas emergentes y adaptar las respuestas con mayor agilidad que los marcos institucionales formales. A través de asociaciones, organi-

zaciones locales y medios de comunicación, se consolidan pautas de uso responsable que complementan la acción legislativa y educativa.

El ámbito familiar configura el núcleo formativo más estable. La mediación activa y el acompañamiento crítico en el hogar influyen de manera decisiva en la construcción de hábitos y en la interpretación que el menor realiza de la autoridad algorítmica. Cuando las familias disponen de competencias digitales suficientes y mantienen una relación de diálogo constante, se reduce el riesgo de uso oculto, dependencia o exposición inadvertida a contenidos problemáticos.

La interdependencia entre estos ámbitos exige una visión sistémica. Un marco normativo avanzado pierde eficacia sin supervisión técnica efectiva; la alfabetización digital se debilita sin coherencia social; y la mediación familiar requiere respaldo institucional para sostenerse. La protección frente a la IAc solo alcanza estabilidad cuando estos niveles operan de forma coordinada.

La IAc puede integrarse en el entorno digital de niños y adolescentes como herramienta de aprendizaje y comunicación, siempre que su despliegue se apoye en estándares técnicos adecuados, marcos normativos exigibles y mediaciones sociales responsables. El reto consiste en equilibrar innovación y protección sin sacrificar el desarrollo integral del menor.

Conclusiones

La IAc plantea a la sociedad contemporánea un desafío que trasciende el plano técnico y jurídico y se proyecta sobre dimensiones normativas, educativas y culturales. En la infancia y la adolescencia, este desafío adquiere especial relevancia, dado que sistemas capaces de simular interacción humana inciden directamente en sujetos que atraviesan procesos de maduración cognitiva, emocional y social. La cuestión central consiste en establecer las condiciones que permitan su integración en el entorno de los menores con plena garantía de sus derechos.

El análisis desarrollado evidencia que los riesgos asociados a la IAc se originan tanto en usos indebidos como en rasgos funcionales propios de su arquitectura. La personalización adaptativa, la memoria conversacional y la simulación de empatía configuran dinámicas de influencia estructural que afectan de manera específica a niños y adolescentes. Los marcos regulatorios vigentes, centrados principalmente en la protección de datos y en la prohibición de contenidos ilícitos, ofrecen una cobertura incompleta frente a estas formas de incidencia.

La protección efectiva requiere una arquitectura multinivel que articule normas jurídicas actualizables, estándares técnicos verificables, mecanismos independientes de supervisión, alfabetización digital crítica y mediaciones familiares y comunitarias coherentes. La debilidad de cualquiera de estos componentes amplía los espacios de vulne-

rabilidad en un entorno tecnológico caracterizado por su rápida evolución.

La protección de la infancia en contextos conversacionales mediados por IA implica asumir una corresponsabilidad sostenida entre instituciones públicas, desarrolladores, educadores, familias, comunidades y los propios menores. Este compromiso compartido orienta el desarrollo y la implementación tecnológica hacia criterios de seguridad, transparencia y respeto al proceso formativo.

La IAc constituye una forma inédita de mediación simbólica que interviene en procesos formativos en curso. Esta circunstancia exige una gobernanza capaz de anticipar las dinámicas de influencia derivadas de su diseño y de establecer límites proporcionados a su capacidad de intervención en el desarrollo del menor. La integración coherente entre técnica, derecho y educación constituye una condición estructural para asegurar que estas tecnologías se orienten de manera estable al desarrollo libre, crítico y seguro de niños y adolescentes, y no interfieran de forma indebida en sus procesos formativos.

Bibliografia

- Ajder, H., Patrini, G., Cavalli, F., Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. Deeptrace.
- Arnd-Caddigan, M. (2015). Sherry Turkle: Alone Together: Why We Expect More from Technology and Less from Each Other. *Clinical Social Work Journal*, 43, 247-248.
- Bai, Y. et al. (2023). *Constitutional AI: Harmlessness from AI Feedback*.
- Craven, S., Brown, S., Gilchrist, E. (2006). Sexual grooming of children: Review of literature and theoretical considerations. *Journal of Sexual Aggression*, 12, 287-299.
- Darling, K. (2017). Who's Johnny? Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy, in Robot Ethics 2.0. En Lin, P., Bekey, G., Abney, K., Jenkins, R. (eds.) (pp. 173-191). Oxford University Press.
- Floridi, L., Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 2020-12, 30 (4), 681-694.
- Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press.
- Livingstone, S., Blum-Ross, A. (2020). *Parenting for a Digital Future: How Hopes and Fears about Technology Shape Children's Lives*. Oxford University Press.

- Livingstone, S., Helsper, E. J. (2007). Gradations in Digital Inclusion: Children, Young People and the Digital Divide. *New Media & Society*, 9(4), 671-696.
- Livingstone, S., Helsper, E. J. (2008). Parental Mediation of Children's Internet Use. *Journal of Broadcasting & Electronic Media*, 52(4), 581-599.
- Mylonas, N., Stylianou, N., Chatzakou, D., Spathi, Th., Alevizos, S., Drandaki, A., Koufakis, A., Kalpakis, G., Tsikrika, T. y Vrochidis, S. (2025). Detección de acoso sexual infantil en línea: Desafíos y futuras direcciones. En Gkotsis, I., Kavallieros, D., Stoianov, N., Vrochidis, S., Diagourtas, D., Akhgar, B. (eds.) *Paradigmas en el desarrollo tecnológico para profesionales de la seguridad. Informática de seguridad y aplicación de la ley*. Springer, Cham.
- Nass, C., Yen, C. (2010). *The Man Who Lied to His Laptop: What Machines Teach Us About Human Relationships*. Penguin Press.
- Steinberg, L. (2008). A Social Neuroscience Perspective on Adolescent Risk-Taking. *Developmental Review*, 28(1), 78-106.
- Turow, J. (2011). *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. Yale University Press.
- Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.

- (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Veale, M., Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97-112.
- Vogt, M., Leser, U., Akbik, A. (2021). *Early detection of sexual predators in chats. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Long Papers) (Vol. 1).*
- Wachter, S., Mittelstadt, B., Russell, C. (2021). *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI.* SSRN Electronic Journal.
- Wallace, R. (2003). *The Elements of AIML Style.* ALICE AI Foundation.
- Wardle, C., Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making.* Council of Europe.
- Weizenbaum, J. (1966). “ELIZA”—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9, 1, 36-45.
- Willison, S. (2022). *Prompt Injection Attacks Against GPT-3.* Simon Willison’s Blog.
- Yeung, K. (2017). Algorithmic Regulation: A Critical Interrogation. *Regulation & Governance*, 12(4), 505-523.

Transit

Explorando _ Dialogando _ Construyendo

Vulnerabilidad. Derechos Humanos y Personas mayores

Jorge Gracia Ibáñez

Tecnologia, occupazione e sostenibilità. La tempesta perfetta?

Nicola Costantino

“Inteligencia Artificial conversacional” en la infancia y la adolescencia

Francisco J. Arteaga Moreno – Anna I. Peirats Navarro

Francisco Javier Arteaga Moreno

Es profesor universitario ordinario, experto en estadística. Ha sido Vicerrector de Investigación en la Universidad Católica de Valencia San Vicente Mártir y director de la Escuela de Doctorado. Es autor de numerosas publicaciones en Estadística, además de transferencia del conocimiento, innovación docente y evaluación de la investigación. Ha participado en proyectos de investigación en Estadística industrial, Economía y Sociología. Forma parte de la Comisión Académica del Programa de Doctorado conjunto en Patrimonio cultural y digitalización entre la Universidad Católica de Valencia y la Universidad Católica de Ávila.

Anna Isabel Peirats Navarro

Es profesora catedrática e investigadora en la Universidad Católica de Valencia San Vicente Mártir y directora del Instituto IVEMIR (Institut Isabel de Villena d'Estudis Medievals i Renaixentistes). Especialista en literatura medieval y renacentista, ética de la IA y humanidades digitales. Su investigación combina la perspectiva humanística con el análisis crítico de las tecnologías emergentes, abordando de manera particular sus implicaciones éticas y sociales. A lo largo de su carrera, ha promovido la integración de enfoques tradicionales y digitales en los estudios literarios. Forma parte de la Comisión Académica del Programa de Doctorado conjunto en Patrimonio cultural y digitalización entre la Universidad Católica de Valencia y la Universidad Católica de Ávila.



“Inteligencia Artificial conversacional”
en la infancia y la adolescencia

Francisco Javier Arteaga Moreno
Anna Isabel Peirats Navarro

ISBN: 978-88-98717-14-9

MADE IN EUROPE 2026 BY EURI CONV

Francisco J. Arteaga Moreno — Anna I. Peirats Navarro

“Inteligencia Artificial conversacional”

en la infancia y la adolescencia

Esta monografía analiza en profundidad los retos y oportunidades que plantea la “Inteligencia Artificial conversacional” (IAC) en interacción con la infancia y la adolescencia. En la primera parte se establece un marco conceptual, ético y técnico, se describe el funcionamiento de los sistemas conversacionales y se examinan los riesgos asociados a su uso por parte de menores, entre los que destacan la exposición a contenidos nocivos, la elaboración de perfiles sin consentimiento, la manipulación ideológica y la generación de dependencia emocional. Asimismo, se identifican vacíos regulatorios, deficiencias en los entornos educativos, así como en la supervisión familiar y comunitaria.

La segunda parte desarrolla un estudio comparado de marcos regulatorios en Europa, Estados Unidos, China y Chile, que pone de relieve diferencias sustanciales en las estrategias de protección infantil, así como vacíos comunes en materia de verificación de edad y en la implementación de auditorías y se describen propuestas de mejora en cinco ámbitos clave: legislativo, educativo, social-comunitario, familiar y de supervisión/auditoría.

Se sostiene que la protección de los menores en la era de la IAC exige un marco de acción integral y flexible, capaz de combinar innovación tecnológica con salvaguardas éticas, que fortalecen tanto las competencias críticas de los menores como la responsabilidad compartida de los diversos actores implicados. Esta monografía se presenta como una contribución al debate internacional sobre cómo garantizar un entorno digital seguro para la infancia, sin menoscabar el potencial educativo y social de la IAC.

worldwide distribution

ITALIAN EDITION

Ebook

www.euriconv.eu



PVP 10€

ISBN 978-88-98717-14-9