**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

# Mecanismos de protección de menores en plataformas digitales y propuesta del prototipo GuardaMirim

*Mecanismos de proteção de menores em plataformas digitais e proposta do protótipo GuardaMirim*

*Protection mechanisms for minors on digital platforms and proposal of the GuardaMirim prototype*

Elis do Nascimento Barbosa – Faculdade Metropolitana de Manaus, elis.nbarbosa22@gmail.com
Henrique da Silva Lima – Faculdade Metropolitana de Manaus, silvalimahenrique14@gmail.com
Laina Santos Almeida – Faculdade Metropolitana de Manaus, lainasantos352@gmail.com
Luiza Natasha Tavares Azevedo – Faculdade Metropolitana de Manaus, luizanatashaxd@gmail.com
Vitor Hugo Guimarães Itou – Faculdade Metropolitana de Manaus, vitoritouguimaraes@gmail.com

**Resumen:**
El acceso a Internet entre niños y adolescentes está creciendo en Brasil, aumentando su exposición a riesgos como el ciberacoso, la manipulación, la coerción y el contacto con extraños. Este artículo analiza los mecanismos tecnológicos para proteger a esta audiencia en entornos digitales, considerando la Ley General de Protección de Datos (LGPD) y el Estatuto Digital de la Niñez y la Adolescencia. (ECA Digital, Ley n.º 15.211/2025). Es un estudio aplicado con un enfoque cualitativo, exploratorio y descriptivo, basado en la investigación bibliográfica y documental y en un análisis comparativo de plataformas (Discord, Instagram, TikTok y WhatsApp) según cinco criterios: verificación de edad, moderación de contenido, controles parentales, privacidad de contenido de menores, y respuesta a incidentes y transparencia. El análisis indicó que los controles parentales y las canales de denuncia se describen en fuentes oficiales para las cuatro plataformas, mientras que la verificación de la edad y la transparencia aparecen de manera desigual en las fuentes analizadas. A partir de este marco, se propone el prototipo de interfaz de la aplicación GuardaMirim para alertar a los tutores sobre signos de comportamiento de riesgo sin exponer el contenido de las conversaciones, en línea con los principios de seguridad, privacidad y el interés superior de los niños y adolescentes. El prototipo utiliza datos simulados, no implementa un clasificador y no ha sido validado con los usuarios. El estudio concluye que la supervisión parental puede fortalecerse con soluciones que complementen las herramientas nativas de las plataformas, que dependen de la validación futura, sin reemplazar la responsabilidad de las plataformas, los tutores y las autoridades.

**Palabras clave:**
Cyberbullying. Privacidad. Supervisión. LGPD. Adolescentes

**Resumo:**
O acesso de crianças e adolescentes à internet cresce no Brasil e amplia a exposição a riscos como cyberbullying, manipulação, coerção e contato com desconhecidos. Este artigo analisa mecanismos tecnológicos de proteção desse público em ambientes digitais, considerando a Lei Geral de Proteção de Dados (LGPD) e o Estatuto Digital da Criança e do Adolescente (ECA Digital, Lei nº 15.211/2025). Trata-se de uma pesquisa aplicada, de abordagem qualitativa e caráter exploratório e descritivo, baseada em pesquisa bibliográfica e documental e em análise comparativa de plataformas (Discord, Instagram, TikTok e WhatsApp), segundo cinco

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

critérios: verificação de idade, moderação de conteúdo, controle parental, privacidade do conteúdo do menor e resposta a incidentes e transparência. A análise indicou que o controle parental e os canais de denúncia estão descritos em fonte oficial nas quatro plataformas, ao passo que a verificação de idade e a transparência aparecem de forma desigual nas fontes analisadas. A partir desse referencial, propõe-se o protótipo de interface do aplicativo GuardaMirim, concebido para alertar os responsáveis sobre sinais de comportamentos de risco sem expor o conteúdo das conversas, em consonância com os princípios de segurança, privacidade e melhor interesse da criança e do adolescente. O protótipo usa dados simulados, não implementa classificador e não foi validado com usuários. Conclui-se que a supervisão parental pode vir a ser fortalecida por soluções complementares às ferramentas nativas das plataformas, o que depende de validação futura, sem substituir a responsabilidade das plataformas, dos responsáveis e das autoridades.

**Palavras-chave:**
Cyberbullying. Privacidade. Supervisão. LGPD. Adolescentes.

**Abstract:**
Internet access among children and adolescents is growing in Brazil, increasing their exposure to risks such as cyberbullying, manipulation, coercion and contact with strangers. This article analyzes technological mechanisms for protecting this audience in digital environments, considering the Brazilian General Data Protection Law (LGPD) and the Digital Statute of Children and Adolescents (ECA Digital, Law No. 15,211/2025). It is an applied study with a qualitative, exploratory and descriptive approach, based on bibliographic and documentary research and on a comparative analysis of platforms (Discord, Instagram, TikTok and WhatsApp) according to five criteria: age verification, content moderation, parental controls, privacy of minors' content, and incident response and transparency. The analysis indicated that parental controls and reporting tools are described in official sources for all four platforms, whereas age verification and transparency appear unevenly in the analyzed sources. Based on this framework, the interface prototype of the GuardaMirim app is proposed to alert guardians about signs of risky behavior without exposing the content of conversations, in line with the principles of security, privacy and the best interests of children and adolescents. The prototype uses simulated data, does not implement a classifier, and has not been validated with users. The study concludes that parental supervision may be strengthened by solutions that complement the platforms' native tools, which depends on future validation, without replacing the responsibility of platforms, guardians and authorities.

**Keywords:**
Cyberbullying. Privacy. Supervision. LGPD. Adolescents.

## 1 Introducción

En el escenario contemporáneo, el avance tecnológico ha hecho que el acceso a Internet por parte de niños y adolescentes sea cada vez más común, modificando la forma en que este público se comunica e interactúa con el mundo. Según la investigación TIC Kids Online Brasil 2025, llevada a cabo por el Centro de Información y Coordinación del Punto BR (NIC.br), el 92% de los niños y adolescentes brasileños entre 9 y 17 años eran usuarios de Internet,

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

evidenciando la presencia de tecnologías digitales en su vida cotidiana. (NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR, 2026).

A pesar de los beneficios que brinda la conectividad, la creciente presencia de esta audiencia en el entorno digital también amplía su exposición a riesgos como el ciberacoso, el contacto con incógnitas, la manipulación, la coerción y el contenido inadecuado. Ante este escenario, la protección de los niños, niñas y adolescentes requiere de un monitoreo familiar y de mecanismos tecnológicos capaces de asistir en la identificación y prevención de conductas nocivas.

En este contexto, la protección de la infancia y la juventud debe tener en cuenta las disposiciones legales relacionadas con la seguridad y la privacidad. La Ley General de Protección de Datos (LGPD), Ley no 13.709/2018, establece que el tratamiento de datos de niños y adolescentes debe observar su interés superior. Además, el Estatuto Digital de la Niñez y la Adolescencia (ECA Digital), establecido por la Ley No. 15.211/2025, establece medidas de protección y responsabilidades para los agentes involucrados en la prestación de servicios tecnológicos.

Aunque plataformas como Discord, Instagram, TikTok y WhatsApp proporcionan mecanismos de seguridad y supervisión parental, estas herramientas presentan limitaciones, especialmente en la identificación de riesgos en conversaciones privadas. Teniendo en cuenta estas necesidades, la investigación propone la aplicación Guarda Mirim, una herramienta de protección complementaria con mecanismos de detección y alerta previstos, que busca preservar la privacidad de los menores sin exponer el contenido original de las conversaciones. En esta etapa, la propuesta se presenta a través de un prototipo navegable con datos simulados.

Ante este escenario, se establece el siguiente problema de investigación: ¿cómo los mecanismos tecnológicos de detección y prevención pueden contribuir a la protección de niños, niñas y adolescentes frente al ciberacoso y otras conductas de riesgo en el entorno digital, considerando LGPD, ECA Digital y la participación de los responsables de la supervisión?

El objetivo general de esta investigación es analizar mecanismos tecnológicos de protección de niños, niñas y adolescentes en entornos digitales, con énfasis en la prevención del ciberacoso y otras conductas de riesgo, considerando la legislación brasileña y proponiendo la aplicación de Guarda Mirim como herramienta complementaria para la supervisión parental.

Para lograr este objetivo, se establecieron los siguientes objetivos específicos: identificar los principales riesgos y formas de violencia digital contra niños, niñas y adolescentes; analizar los mecanismos de protección de las plataformas Discord, Instagram,

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

TikTok y WhatsApp, considerando los requisitos de LGPD y ECA Digital; y presentar los requisitos, la arquitectura propuesta y el prototipo de interfaz de Guarda Mirim, respetando los principios de seguridad y privacidad.

Se plantea la hipótesis de que los mecanismos tecnológicos de detección y prevención, desarrollados de acuerdo con los principios de seguridad, privacidad y mejor interés de los niños y adolescentes, pueden contribuir a la identificación temprana de situaciones de riesgo y fortalecer el desempeño de los responsables, sin reemplazar el monitoreo familiar.

En cuanto a los procedimientos metodológicos, la investigación se caracteriza por ser aplicada, cualitativa, exploratoria y descriptiva, basada en la investigación bibliográfica y documental y en el análisis comparativo de los mecanismos de protección de cuatro plataformas digitales. El desarrollo del prototipo se guió por el enfoque de investigación de la ciencia del diseño, contemplando la definición de los requisitos, el modelado y la implementación de las interfaces.

Como resultado, se identificaron limitaciones en los mecanismos de protección de las plataformas y se desarrolló un prototipo navegable de Guarda Mirim, compuesto por siete pantallas. Sin embargo, la detección automatizada aún no se ha implementado, y se necesitan evaluaciones futuras para verificar su efectividad.

Finalmente, el artículo está organizado en cinco secciones, además de esta introducción: marco teórico, metodología, desarrollo del prototipo de Guarda Mirim, resultados y discusión, y consideraciones finales.

## 2 Marco teórico

### 2.1 Riesgos digitales y violencia contra niños, niñas y adolescentes

En 2025, el 92% de las personas de 9 a 17 años en Brasil eran usuarios de Internet, alrededor de 24 millones de niños y adolescentes. (NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR, 2026). Para esta audiencia, Internet es un espacio de estudio, ocio y socialización, pero también de exposición a riesgos que dependen no solo del contenido al que se accede, sino de quién entra en contacto con el menor y cómo se desarrollan las relaciones en línea.

El ciberacoso es una forma de violencia digital. De conformidad con la Ley No. 13.185/2015, “existe intimidación sistemática en la red informática mundial (cyberbullying), al utilizar los instrumentos adecuados para depreciar, incitar a la violencia, adulterar fotos y datos

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

personales con el fin de crear medios de restricciones psicosociales". (BRASIL, 2015, art. 2º, parágrafo único). Ley No. 14,811/2024 comenzó a tratar la intimidación sistemática, incluida la intimidación virtual, como un delito. (BRASIL, 2024). En comparación con el acoso cara a cara, el ciberacoso puede llegar a audiencias ampliadas, mantener registros accesibles y ocurrir bajo anonimato, características relevantes para su detección y afrontamiento. (SALAWU; HE; LUMSDEN, 2020).

Para los fines de esta investigación, los riesgos se organizaron en cuatro categorías, que también guían el prototipo de GuardaMirim:

- Ciberacoso: agresiones repetidas, humillaciones y exclusión de un menor a través de mensajes, publicaciones o grupos.
- Lenguaje ofensivo: maldiciones, discursos de odio y amenazas, que pueden ocurrir de forma aislada, sin la típica repetición del ciberacoso, pero con potencial de daño.
- Imagen inadecuada: recibir o compartir imágenes inadecuadas para la edad, como la desnudez y la violencia, incluidas imágenes íntimas del propio menor.
- Contacto desconocido: acercamiento de adultos o desconocidos que puede evolucionar hacia la captación.

La manipulación y la coerción no son necesariamente categorías aisladas, sino estrategias transversales que pueden estar presentes en diferentes manifestaciones de violencia. Los estudios sobre la mediación en línea de niños y adolescentes y sobre el control coercitivo señalan que estas prácticas pueden desarrollarse progresivamente, a través de la construcción de la confianza, el control emocional y el aislamiento social de la víctima, estableciendo condiciones que favorecen el dominio ejercido por el agresor. (WHITTLE et al., 2013; WIENER, 2017; MCLINDON et al., 2025). Este patrón aparece, por ejemplo, en la investigación policial denunciada en 2026, en la que los sospechosos, tras ganarse la confianza de niños y adolescentes, nos someten a coerción, aislamiento y violencia psicológica. (FANTTI, 2026). Las categorías tampoco son excluyentes, porque la misma situación puede combinar, por ejemplo, contacto desconocido, imagen inadecuada y coerción.

## 2.2 Marco jurídico para la protección de los niños, niñas y adolescentes en el entorno digital

En Brasil, la Ley General de Protección de Datos Personales (LGPD) dedica el art. 14 al tratamiento de datos de niños y adolescentes, que debe realizarse en su mejor interés (BRASIL, 2018). En el caso de los niños, se requiere un consentimiento específico y se hace

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

hincapié en al menos uno de los padres o tutor legal, y la información sobre el tratamiento debe presentarse en un lenguaje sencillo y accesible. (BRASIL, 2018, art. 14, §§ 1º e 6º). Los principios del Art. 6a, en particular, la necesidad, la transparencia, la seguridad y la prevención. Para las soluciones tecnológicas, esto significa recopilar solo los datos estrictamente necesarios y explicar claramente qué se trata.

El Estatuto Digital de la Niñez y la Adolescencia (ECA Digital), establecido por la Ley No. 15.211/2025, entró en vigor el 17 de marzo de 2026 (AGÊNCIA BRASIL, 2026). Se aplica a los productos y servicios de tecnología de la información dirigidos a niños, niñas y adolescentes o que puedan tener acceso a ellos, que abarca las redes sociales, las aplicaciones de mensajería y las plataformas de juegos. (BRASIL, 2025; FUNDAÇÃO ABRINQ PELOS DIREITOS DA CRIANÇA E DO ADOLESCENTE, 2026). Entre sus principales requisitos se encuentran:

- Verificación de la edad por mecanismos fiables, siendo insuficiente la mera autodeclaración;
- Supervisión parental, con herramientas accesibles para que los padres y tutores puedan seguir y configurar el uso, incluido el archivo adjunto de las cuentas de adolescentes de hasta 16 años a la de un tutor;
- Transparencia, con informes periódicos de las plataformas (art. 31), cuyo primer período de publicación fue fijado por la Agencia Nacional de Protección de Datos el 17 de septiembre de 2026 (AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS, 2026).

La ley también establece que el cumplimiento de estas obligaciones no puede dar lugar a mecanismos de vigilancia masiva, genérica o indiscriminada. (BRASIL, 2025, art. 34, § 1º). Esta predicción es fundamental para este estudio, porque delimita lo que una solución de protección puede hacer sin violar la privacidad del menor.

En el escenario internacional, la Regla de Protección de la Privacidad Infantil en Línea de los Estados Unidos (COPPA) requiere el consentimiento verificable de los padres para la recopilación de datos de niños menores de 13 años. (FEDERAL TRADE COMMISSION, 2025), y el art. El 8 de Reglamento General de Protección de Datos de la Unión Europea (GDPR) fijó la edad de consentimiento en 16 años, y los Estados miembros pueden reducirla hasta 13 años. (UNIÃO EUROPEIA, 2016). Estos hitos se centran en el consentimiento y el tratamiento de datos, mientras que ECA Digital también se ocupa del diseño de servicios, la verificación de la edad y la supervisión de los padres.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

### 2.3 Mecanismos de protección en entornos digitales

En el marco jurídico se identifican cinco mecanismos de protección que corresponden a los criterios del cuadro 1.

Verificación de edad: busca establecer el grupo de edad del usuario para ofrecer una experiencia adecuada y restringir el contenido inapropiado. Difiere de la verificación de identidad, ya que identifica si el usuario es más pequeño, y no quién es. La autodeclaración es considerada insuficiente por la ECA Digital (BRASIL, 2025).

Modulación de contenido: incluye detección automatizada, revisión humana y canales de informes utilizados para identificar y eliminar comportamientos dañinos. La literatura sobre la detección automatizada del ciberacoso apunta a desafíos de interpretación y clasificación contextuales (SALAWU; HE; LUMSDEN, 2020); estos sistemas pueden equivocarse y, en servicios con criptografía de extremo a extremo, el análisis del contenido por parte de la plataforma está técnicamente restringido; por lo tanto, la moderación aislada no es suficiente para proteger a los menores en conversaciones privadas.

Control parental: consiste en herramientas que permiten al responsable vincular cuentas, establecer plazos de uso y contactos, además de monitorizar ciertos aspectos del uso de los servicios. El ECA Digital establece la necesidad de disponibilidad y accesibilidad de estas herramientas (BRASIL, 2025). Sin embargo, estudios sobre el control parental identifican tensiones entre protección, autonomía y privacidad de los adolescentes (WISNIEWSKI et al., 2017; GHOSH et al., 2018a).

Privacidad del contenido del menor: se refiere al grado de acceso que la plataforma, el responsable o terceros tienen al contenido de las interacciones de los menores. Este mecanismo está relacionado con los principios del interés superior y la necesidad. (BRASIL, 2018), así como la prohibición de la vigilancia masiva proporcionada por la Digital ECA (BRASIL, 2025, art. 34, § 1º). Por lo tanto, la supervisión debe buscar el nivel más bajo posible de exposición de datos, sin comprometer la protección necesaria.

Respuesta a incidentes y transparencia: abarca la publicación de informes, la asistencia a las denuncias y la presentación de plataformas a la inspección y las sanciones previstas en la ley. Estos procedimientos permiten a la sociedad y al gobierno evaluar, basándose en documentos públicos, si los mecanismos de protección son efectivos en la práctica. (BRASIL, 2025, art. 31).

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

### 2.4 Síntesis y brecha

La literatura científica y la legislación examinada apuntan a dos tendencias. (WISNIEWSKI et al., 2017; GHOSH et al., 2018a; BRASIL, 2025). El primero es la rendición de cuentas de las plataformas para verificar la edad, moderar el contenido y ofrecer la supervisión de los padres. La segunda es la preocupación por la privacidad del menor, que limita el acceso sin restricciones al contenido de las conversaciones. Los mecanismos nativos, sin embargo, operan dentro de cada servicio, lo que obliga a los responsables a seguir varios paneles, y no siempre logran alertarlo a tiempo sobre situaciones como el ciberacoso, el aseo personal y la coerción, que se manifiestan en conversaciones privadas.

Esta brecha indica la necesidad de soluciones complementarias que identifiquen signos de riesgo e informen al responsable sin entregar el contenido de las conversaciones, en línea con los principios de minimización y de mejor interés. Es en este espacio donde se inserta la propuesta de GuardaMirim, detallada en el apartado 4, y cuya adecuación a la marca aquí presentada se discute a continuación.

## 3 Material y método

### 3.1 Caracterización de la investigación

Esta investigación es de carácter aplicado, con un enfoque cualitativo y un carácter exploratorio y descriptivo. (GIL, 2017). Se realizó en dos frentes complementarios. El primero es un análisis documental comparativo de los mecanismos de seguridad y moderación adoptados por las plataformas digitales para proteger a los niños, niñas y adolescentes. El segundo es la construcción de un artefacto tecnológico, el prototipo GuardaMirim, guiado por la investigación en ciencias del diseño, en el que el desarrollo de un artefacto forma parte del método de investigación. (HEVNER et al., 2004; PEFFERS et al., 2007).

La opción de análisis documental se justifica por el carácter reciente del objeto: Ley No. 15.211/2025 (ECA Digital) entró en vigor el 17 de marzo de 2026 (AGÊNCIA BRASIL, 2026), y las plataformas todavía ajustan sus recursos y documentos. Como los materiales difieren en alcance y nivel de detalle, la comparación no presupone una equivalencia directa entre ellos. La investigación no incluyó entrevistas ni recolección directa de datos con las empresas, restringida al análisis de documentos públicos y al desarrollo del prototipo.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

### 3.2 Fuentes de datos y criterios de selección

Se consultaron tres categorías de fuentes: i) legislación y documentos oficiales, incluida la Ley No. 15.211/2025 (BRASIL, 2025), Ley no 13.709/2018 (BRASIL, 2018), la regla norteamericana COPPA (FEDERAL TRADE COMMISSION, 2025), el Reglamento General de Protección de Datos de la Unión Europea (UNIÃO EUROPEIA, 2016) y materiales de la Agencia Nacional de Protección de Datos (ANPD); (ii) documentos institucionales de cada plataforma, seleccionados entre seguridad, ayuda, privacidad, supervisión parental y transparencia de la Discord, Instagram/Meta, TikTok y WhatsApp (DISCORD, [2026?]; META, 2026a, 2026b, 2026c; TIKTOK, [2026?], 2026; WHATSAPP, [2026?]); y (iii) materiales institucionales, informes de investigación e informes sobre inspección, litigios, protección de niños y jóvenes y cambios en los recursos de la plataforma.

Las fuentes primarias fueron consultadas en los portales oficiales de Planalto, ANPD, Comisión Federal de Comercio y EUR-Lex, así como en los centros de seguridad y ayuda de las plataformas. Para localizar estudios y noticias, se utilizaron Google Search, Google Scholar, SciELO y CAPES Journal Portal, con términos como "ECA Digital", "verificación de edad", "control parental" y "ciberacoso", combinados por operadores AND y OR; el análisis se basó en los documentos recuperados y verificados.

El corpus final estuvo compuesto por 20 documentos, contabilizados individualmente: cuatro marcos normativos (ECA Digital, LGPD, COPPA y GDPR), siete materiales institucionales de las plataformas (uno de Discord, dos de Meta/Instagram, dos de WhatsApp y dos de TikTok), una página oficial de la ANPD sobre el ECA Digital, dos materiales institucionales/de investigación (Fundação Abrinq y NIC.br/Cetic.br) y seis reportajes (Agência Brasil, Migalhas, Piauí Hoje, Showmetech, Canaltech y Jornal de Brasília/Folhapress). Las referencias teóricas y metodológicas utilizadas únicamente para fundamentar la investigación no se incluyeron en este recuento.

Los criterios de inclusión fueron: (a) publicación o actualización entre 2025 y 2026 (para páginas sin fecha explícita, se consideró la versión disponible en la colección, con registro de la fecha de acceso), a excepción de marcos normativos anteriores (LGPD, COPPA y GDPR); (b) relación directa con al menos uno de los criterios de la sección 3.4; e (c) identificación de la autoría o el organismo responsable. Se dio prioridad a las fuentes primarias; los informes se utilizaron de manera complementaria, como informes periodísticos y no como hechos

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

auditados. La colección principal tuvo lugar en septiembre de 2026, y la conferencia final de los enlaces se realizó hasta el 1 de noviembre de 2026.

### 3.3 Selección de plataformas

Se analizaron cuatro plataformas: Discord, Instagram, TikTok y WhatsApp. El recorte se basó en (i) su relevancia en el debate público y regulatorio brasileño en 2026, incluida la acción civil pública de la Fiscalía General de la Unión contra la Discord. (MIGALHAS, 2026), la investigación de la Fiscalía Federal sobre la Seguridad de la Discord (PIAUÍ HOJE, 2026) y la alteración de la clasificación indicativa (SHOWMETECH, 2026); y (ii) su marco en la ley, que abarca las redes sociales, las aplicaciones de mensajería y las plataformas de juegos (FUNDAÇÃO ABRINQ PELOS DIREITOS DA CRIANÇA E DO ADOLESCENTE, 2026). Es un muestreo intencional, sin pretensión de representatividad, y las diferencias entre los tipos de servicio limitan la comparabilidad directa.

### 3.4 Criterios de análisis

La comparación se estructuró en cinco criterios, derivados del marco teórico (Quadro 1).

**Cuadro 1 – Criterios de análisis y pregunta guía**

| Criterio | Pregunta orientadora |
|---|---|
| Verificación de edad | ¿La plataforma adopta algún mecanismo además de la autodeclaración? |
| Moderación de contenido | ¿Hay moderación automatizada, revisión humana y un canal de denuncia? ¿Qué límites aparecen en las fuentes? |
| Control parental | ¿Existe supervisión parental nativa? ¿Qué puede ver y configurar el responsable? |
| Privacidad del contenido del menor | ¿El responsable o la plataforma acceden al contenido de las interacciones? |
| Respuesta a incidentes y transparencia | ¿Hay informes, sanciones o actuaciones regulatorias públicas? |

*Fuente: Elaborado por los autores (2026).*

### 3.5 Procedimientos

El procedimiento siguió cuatro etapas: (1) revisión bibliográfica y normativa y definición de los criterios; (2) recogida y organización de la información en una matriz plataforma × criterio; (3) comparación de los mecanismos identificados con las exigencias del

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

ECA Digital y la LGPD; y (4) interpretación mediante análisis de contenido categorial (BARDIN, 2016), con los cinco criterios del Cuadro 1 como categorías. Cada celda se clasificó como mecanismo identificado en fuente oficial, identificado únicamente en fuente periodística o no identificado en las fuentes consultadas, lo que indica ausencia de información pública y no ausencia del recurso.

### 3.6 Aspectos éticos y legales

La investigación utilizó exclusivamente documentos de acceso público y no involucró a seres humanos. Por ello, no fue necesaria la aprobación de un comité de ética, de conformidad con el artículo 1, párrafo único, inciso II, de la Resolución CNS n.º 510/2016 (BRASIL, 2016). No se recogieron datos personales de niños o adolescentes: las alertas, los nombres y las métricas del prototipo son ficticios. El diseño del artefacto consideró el interés superior del niño y la minimización de datos (BRASIL, 2018, art. 14), así como la prohibición de vigilancia masiva (BRASIL, 2025, art. 34, § 1º). Cualquier estudio futuro con participantes requerirá evaluación ética previa al reclutamiento.

## 4 Desarrollo del prototipo de GuardaMirim

### 4.1 Visión general y alcance

GuardaMirim es una aplicación para apoyar la protección de niños y jóvenes, diseñada para detectar señales de acoso cibernético y otros riesgos digitales y alertar a la persona responsable sin exponer el contenido de las conversaciones del niño. El desarrollo siguió tres pasos iterativos, típicos de la investigación en ciencias del diseño: identificación de problemas, diseño y desarrollo, y evaluación, que resumen las seis actividades del modelo original. (PEFFERS et al., 2007).

El alcance del artefacto está limitado, en esta etapa, al diseño de la interfaz y al flujo de navegación. El motor de clasificación de contenido no se implementó y todos los datos mostrados se simulan. La arquitectura completa, descrita en la sección 4.3, es una propuesta para futuros pasos.

### 4.2 Requerimientos

Los requisitos se derivaron de los criterios de análisis (Quadro 1) y las restricciones legales discutidas en el marco teórico. El gráfico 2 presenta los principales.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

**Cuadro 2 - Necesidades de GuardaMirim**

| Código | Tipo | Descripción |
|---|---|---|
| RF01 | Funcional | Permitir la selección entre el perfil del menor y el perfil del responsable |
| RF02 | Funcional | Exigir autenticación (PIN de 4 dígitos, utilizado únicamente como barrera mínima en el prototipo) para acceder al perfil del responsable |
| RF03 | Funcional | Mostrar el estado de protección del dispositivo y las incidencias por categoría durante los últimos 7 días |
| RF04 | Funcional | Enumerar alertas recientes con su gravedad (baja, media o alta), aplicación de origen y hora |
| RF05 | Funcional | Detallar la alerta con metadatos y un resumen de la clasificación |
| RF06 | Funcional | Ofrecer al responsable las acciones de marcar como resuelto, consultar orientaciones para conversar y pausar la aplicación de origen durante 1 hora (requisito conceptual, sujeto a los permisos del sistema operativo) |
| RF07 | Funcional | Informar al menor, en un panel propio, de que la protección está activa y de qué se analiza |
| RF08 | Funcional | Permitir al responsable ajustar la sensibilidad de detección y activar o desactivar categorías de alerta |
| RF09 | Funcional | Notificar al responsable nuevas alertas, incluso en la pantalla de bloqueo, sin mostrar el contenido bruto |
| RF10 | Funcional | Permitir al responsable registrar su acuerdo o desacuerdo con la clasificación de la alerta (retroalimentación) |
| RNF01 | No funcional | Privacidad por defecto: ninguna pantalla puede mostrar el contenido bruto de mensajes o imágenes |
| RNF02 | No funcional | Comunicar la gravedad mediante color y texto simultáneamente, con tonos no alarmistas |
| RNF03 | No funcional | Accesibilidad: foco visible, áreas táctiles amplias y respeto por la preferencia de movimiento reducido |

*Fuente: Elaborado por los autores (2026).*

Las categorías de riesgo del prototipo (ciberacoso, lenguaje ofensivo, imagen inadecuada y contacto desconocido) reflejan la taxonomía de riesgo presentada en la sección 2.1.

### 4.3 Arquitectura propuesta

La arquitectura fue diseñada en capas, de modo que el contenido de las interacciones permanezca en el dispositivo del menor. (Figura 1 e Quadro 3). Solo la clasificación del evento (categoría, gravedad, aplicación de origen, contexto y tiempo) se envía al responsable.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

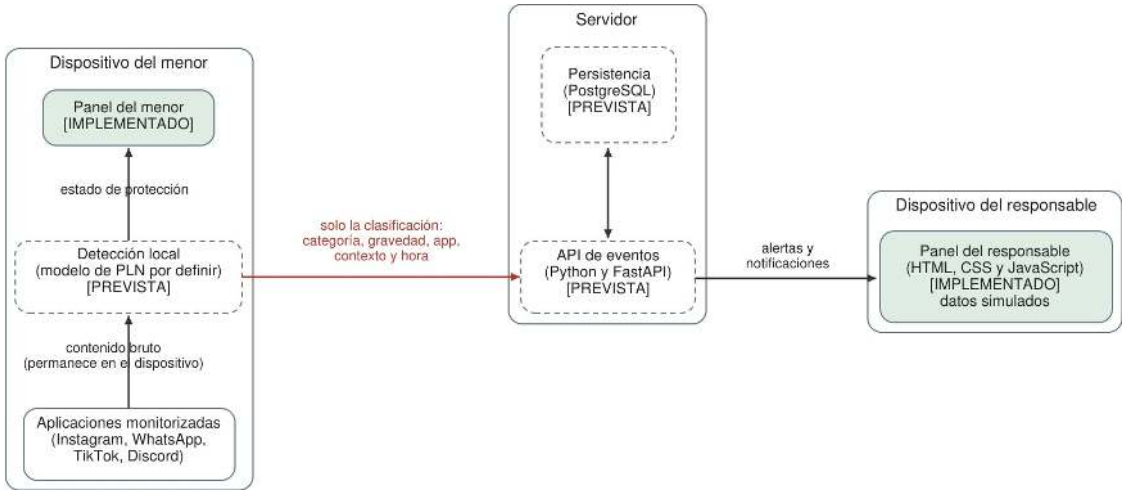
**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

**Cuadro 3 – Capas de arquitectura y situación de implementación**

| Capa | Función | Tecnología | Estado de implementación |
|---|---|---|---|
| Interfaz (perfiles del menor y del responsable) | Pantallas y navegación | HTML, CSS y JavaScript | Implementada (datos simulados) |
| Detección local | Clasificar el contenido en el dispositivo del menor | Modelo de procesamiento del lenguaje natural (PLN) por definir | Prevista |
| API de eventos | Recibir y entregar clasificaciones de eventos | Python y FastAPI | Prevista |
| Persistencia | Almacenar eventos y preferencias | PostgreSQL | Prevista |

*Fuente: Elaborado por los autores (2026).*

**Figura 1 - Arquitectura propuesta de GuardaMirim**




*Fuente: Elaborado por los autores (2026). Los bloques trazados indican componentes predichos, aún no implementados.*

Esta organización busca reunirse con la ECA Digital, que prohíbe mecanismos de vigilancia masiva, genérica o indiscriminada (BRASIL, 2025, art. 34, § 1º). Lo que circula entre las capas es la clasificación del evento, nunca el texto o la imagen que lo originó. Las políticas de retención y eliminación de metadatos de eventos aún deben definirse, con plazos, propósito, control de acceso, exclusión y registro de auditoría.

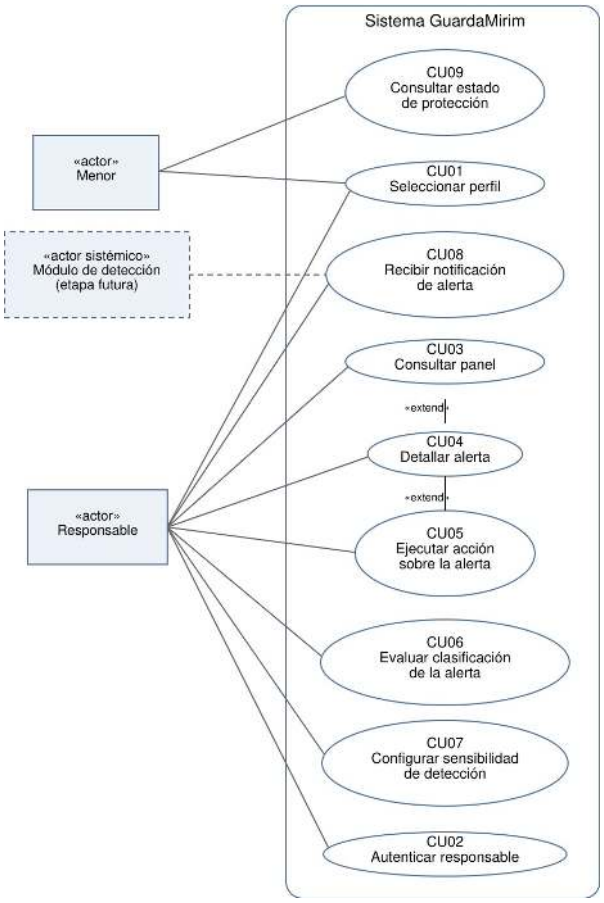
### 4.4 Modelado

El modelado identificó dos actores humanos, el menor y el responsable, y el módulo de detección como un futuro actor sistémico. Se plantearon nueve casos de uso (CU01 a CU09), presentado en el diagrama de la Figura 2.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

**Figura 2 - Diagrama de casos de uso de GuardaMirim**




*Fuente: Elaborado por los autores (2026).*

### 4.5 Implementación e interfaces

El prototipo se implementó en HTML, CSS y JavaScript, como flujo navegable de siete pantallas. El gráfico 4 relaciona cada pantalla con su función y con la decisión de privacidad correspondiente, y la figura 3 reúne las capturas. El flujo del controlador corre a través de los paneles (a) a (f): selección de perfil, autenticación, panel, detalles de alerta, configuración de sensibilidad y notificación en la pantalla de bloqueo; el flujo del menor corresponde al panel (g). El prototipo está disponible a petición de los autores.

**Cuadro 4 – Pantallas de prototipos y decisiones de privacidad**

| Pantalla | Función | Decisión de privacidad |
|---|---|---|
| Selección de perfil (Fig. 3a) | Separar los accesos | Aviso que informa de que el análisis es local |
| Autenticación (Fig. 3b) | Proteger el panel del responsable | Barrera mínima antes de acceder a datos sensibles |

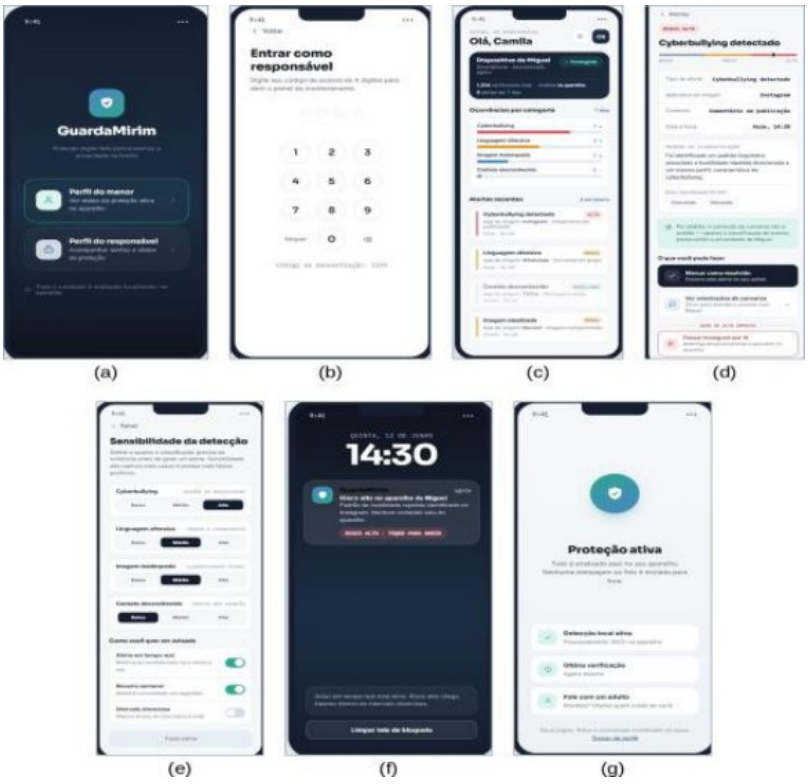
Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

| Pantalla | Función | Decisión de privacidad |
|---|---|---|
| Panel del responsable (Fig. 3c) | Estado, gráfico de categorías y lista de alertas | Muestra únicamente metadatos |
| Detalles de la alerta (Fig. 3d) | Metadatos, resumen y acciones | Aviso explícito de que el contenido no se muestra; acciones dirigidas a la mediación familiar |
| Sensibilidad de detección (Fig. 3e) | Ajustar el rigor de la detección y las categorías monitorizadas | Control del responsable sobre el nivel de monitorización, sin acceso al contenido |
| Notificación (Fig. 3f) | Avisar al responsable de una nueva alerta en la pantalla de bloqueo | Mensaje con categoría y gravedad, sin fragmentos de la conversación |
| Panel del menor (Fig. 3g) | Confirmar la protección activa | Transparencia sobre lo que se analiza |

*Fuente: Elaborado por los autores (2026).*

**Figura 3 - Pantallas del prototipo de GuardaMirim (a a g)**




*Fuente: Elaborado por los autores (2026).*

### 4.6 Evaluación interna de la coherencia de la interfaz

La evaluación de esta etapa fue interna: el equipo revisó la coherencia entre el discurso de privacidad comunicado al usuario y lo que realmente muestra cada pantalla. Esta revisión resultó, por ejemplo, la inclusión de una advertencia de privacidad en la pantalla de detalles de

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

alerta. El prototipo aún no se ha sometido a pruebas de usabilidad o validación con familias, y no hubo medición del rendimiento de la clasificación.

## 5 Resultados y discusión

### 5.1 Comparación de los mecanismos de protección en las plataformas

El Cuadro 5 sintetiza la matriz plataforma × criterio construida según la sección 3.5. Cada celda indica si el mecanismo se identificó en fuente oficial (O), únicamente en fuente periodística (J) o no se identificó en las fuentes consultadas (–). Como se previó en el método, la tercera situación indica ausencia de información en los documentos analizados y no ausencia del recurso en la plataforma.

**Cuadro 5 – Matriz plataforma × criterio: mecanismos identificados en las fuentes analizadas**

| Criterio | Discord | Instagram | TikTok | WhatsApp |
|---|---|---|---|---|
| Verificación de edad | J - registro con fecha de nacimiento; la Unión alega que es insuficiente (MIGALHAS, 2026) | - No identificado en la fuente analizada (META, 2026b) | O – edad mínima de 13 años; formulario de fecha de nacimiento y modelos de detección de cuenta de menores de 13 años (TIKTOK, 2026) | - No identificado; la fuente no describe el método (META, 2026a) |
| Moderación de contenido y denuncias | O - herramientas de denuncia y bloqueo (DISCORD, [2026?]). J – la acción pide moderación en portugués y detección de transmisiones de riesgo (MIGALHAS, 2026) | O – canal de denuncia de contenidos ilegales en Brasil, desde marzo de 2026 (META, 2026b) | O - tecnología y revisión humana; denuncias por cualquier persona; se denuncian casos de explotación sexual infantil (TIKTOK, 2026) | O – denuncia y bloqueo de contactos y mensajes; advertencia de contexto en mensajes de extraños; suspensión por incumplimiento de los Términos de Servicio, que prohíben el contenido de abuso, intimidación o amenaza (WHATSAPP, [2026?]) |
| Control parental | O - Centro Familiar: vinculación de cuentas; el responsable ve amigos, servidores y contactos recientes (DISCORD, [2026?]) | O - supervisión activable sin autorización del adolescente; alertas de búsquedas de suicidio y autolesiones (META, 2026b); gestión centralizada en el Centro Familiar, con una invitación única a Instagram, Facebook, Messenger y Meta Horizon (META, 2026c) | O – Sincronización familiar: límite de tiempo, restricción de contenido y gestión de mensajes (TIKTOK, 2026) | O – cuentas administradas por los padres, para preadolescentes (META, 2026a). J – nuevos controles para adolescentes, anunciados el 30 de septiembre de 2026 (CANALTECH, 2026) |

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

| Criterio | Discord | Instagram | TikTok | WhatsApp |
|---|---|---|---|---|
| Privacidad del contenido del menor | O - el responsable no accede al contenido de los mensajes privados (DISCORD, [2026?]) | – No identificado; la fuente cita solo temas de conversaciones con Meta AI (META, 2026b) | O – cuentas privadas de 13 a 17 años por defecto, con publicaciones visibles solo para seguidores aprobados; mensajes directos no disponibles para usuarios de 13 a 15 años (TIKTOK, [2026?]; TIKTOK, 2026). Las fuentes no aclaran si los padres ven los mensajes. | O – conversaciones con cifrado de extremo a extremo (META, 2026a). J – los padres no ven el contenido de las conversaciones (CANALTECH, 2026) |
| Respuesta a incidentes y transparencia | J – Acción Civil Pública de AGU, con una solicitud de indemnización de al menos R$ 500 millones (MIGALHAS, 2026) | – No identificado; la fuente no menciona informes de transparencia (META, 2026b) | O – informes trimestrales en el Centro de Transparencia; acceso de investigadores a datos públicos de menores (TIKTOK, 2026) | - No identificado en las fuentes analizadas |

Leyenda: O = fuente oficial; J = únicamente fuente periodística; – = no identificado en las fuentes consultadas. Fuente: elaborado por los autores (2026), a partir de los documentos citados en cada celda.

Se observó que el control parental y los canales de denuncia eran los criterios con una fuente oficial de información en las cuatro plataformas. Por el contrario, el control de edad se describió en la fuente oficial solo en TikTok. En el caso de Discord, la única referencia encontrada proviene de una fuente periodística, que informa la afirmación de la Unión de que la declaración de la fecha de nacimiento es insuficiente, y para Instagram y WhatsApp, las fuentes analizadas no detallan el método. En cuanto a los informes de transparencia, solo TikTok los describe en fuente oficial; en Discord, la fuente encontró acuerdos con acciones regulatorias, y no con el informe.

En cuanto a la privacidad, Discord y WhatsApp informan que el responsable no accede al contenido de las conversaciones, y, en WhatsApp, la criptografía de punta a punta impide incluso el acceso de la empresa, según el Objetivo. (META, 2026a). Esta convergencia es compatible con el sellado de la vigilancia masiva (BRASIL, 2025, art. 34, § 1º) y confirma la tensión entre la protección y la privacidad de la sección 2.3: los controles nativos entregan a las configuraciones responsables y la información de uso, pero no el contenido. Las alertas identificadas se refieren a eventos específicos, como búsquedas de suicidio y automutilación en Instagram. (META, 2026b) y la entrada en grupos en WhatsApp (CANALTECH, 2026). En las fuentes consultadas, no se identificaron descripciones públicas de alertas a los responsables sobre signos de acoso cibernético o aseo en conversaciones privadas, una brecha que motiva a GuardaMirim.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

Es importante destacar los límites de la comparación. La matriz depende de los documentos disponibles hasta el 1 de noviembre de 2026, y las plataformas cambian sus recursos con frecuencia; la nota Meta en Instagram, por ejemplo, se actualizó el 29 de octubre de 2026. La información sobre los fallos de Discord se deriva de las afirmaciones de la Unión comunicadas por la prensa, y no de la auditoría.

### 5.2 Limitaciones del prototipo y futuros pasos de validación

Esta subsección presenta las limitaciones del prototipo y los pasos futuros de GuardaMirim. Los resultados aquí descritos no se obtuvieron en esta investigación: son objetivos de desarrollo y evaluación. El prototipo demuestra la interfaz y el flujo de navegación con datos simulados; no implementa un clasificador, no ha sido validado con familias y no demuestra el acceso al contenido de otras aplicaciones. Los pasos futuros son: (i) evaluar la viabilidad de un clasificador en el dispositivo; (ii) estudiar la mejora del modelo con protección de datos locales; (iii) validar la experiencia con tutores y adolescentes; y (iv) especificar la gobernanza de datos. Los criterios numéricos de aprobación deben definirse antes de las pruebas confirmatorias, basándose en un estudio piloto, y no mediante la transposición de los resultados de otros conjuntos de datos o dispositivos.

El primer paso es determinar si es factible reemplazar los datos simulados por un clasificador ejecutado en el dispositivo del usuario. La inferencia local es una propiedad de arquitectura, pero no resuelve cómo la aplicación obtendría el texto a analizar. Será necesaria una prueba de concepto que identifique la fuente del contenido, los permisos requeridos y las restricciones técnicas y de privacidad. Una alternativa de alcance inicial es analizar el contenido que el propio usuario comparte voluntariamente con GuardaMirim, sin asumir la lectura sin restricciones de otras aplicaciones.

Para el portugués brasileño, hay modelos y conjuntos de datos que pueden apoyar la investigación: BERTimbau, un modelo pre-entrenado (SOUZA; NOGUEIRA; LOTUFO, 2020); ToLD-Br, de publicaciones de lenguaje tóxico, cuyo trabajo de referencia reporta alrededor del 76% de macro-F1 en tarea binaria (LEITE et al., 2020); y HateBR, comentarios de Instagram notados por ofensiva y discurso de odio (VARGAS et al., 2022). Estos números sirven como referencia para comparar resultados publicados, y no como metatransferibles a GuardaMirim.

La evaluación debe especificar primero qué se clasificará, ya que el lenguaje ofensivo, el discurso de odio y el ciberacoso están relacionados pero no son fenómenos equivalentes.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

Exigirá categorías operativas, directrices de anotación, datos de prueba separados de la capacitación y la presentación de informes de macro-F1, precisión y recuperación por categoría, con examen de falsos positivos y negativos. La escasez de bases representativas etiquetadas es uno de los desafíos centrales del área. (SALAWU; HE; LUMSDEN, 2020), y la jerga y la ironía de los adolescentes requerirán un protocolo ético adecuado. La ejecución en dispositivos móviles es otra prueba: MobileBERT redujo el tamaño y la latencia en puntos de referencia específicos (SUN et al., 2020), que no predice el rendimiento de un modelo en portugués en el público objetivo. Debe medirse el tamaño, la latencia, la memoria y el consumo de energía, con límites fijados después del estudio piloto.

Otro paso es evaluar si los botones "Concordo" y "Discordo" de la pantalla de detalles de alerta (RF10) puede generar retroalimentación útil para mejorar el clasificador sin enviar el mensaje original a un servidor. Esta retroalimentación no es equivalente a una etiqueta de verdad: ‘Discordo’ indica desacuerdo, pero no prueba un falso positivo. El aprendizaje federado, en el que el entrenamiento ocurre en los dispositivos y solo se agregan las actualizaciones de parámetros (MCMAHAN et al., 2017), es una alternativa para investigar; ya se ha aplicado a la previsión de texto en teclados móviles (HARD et al., 2018), sin demostrar su idoneidad para GuardaMirim. Las actualizaciones del modelo también pueden revelar información indirecta, por lo que "no enviar el texto" no garantiza la privacidad absoluta.

La tercera etapa es evaluar la usabilidad y comprensión del prototipo con tutores y adolescentes (seção 4.6). Los estudios sobre aplicaciones de control parental apuntan a la tensión entre la supervisión, la seguridad, la autonomía y la privacidad de los adolescentes (WISNIEWSKI et al., 2017; GHOSH et al., 2018a), y otro estudio asoció el uso de aplicaciones de monitoreo con más experiencias de riesgo en línea, mientras que la participación directa de los responsables se asoció con menos problemas; Debido a que es evidencia contextual, no debe tomarse como causal o generalizada (GHOSH et al., 2018b). Se espera un estudio piloto con los dos perfiles, combinando tareas observadas, entrevistas semiestructuradas y la Escala de Usabilidad del Sistema (SUS). (BROOKE, 1996). El promedio de alrededor de 70 puntos y el valor de 71,4 asociados con el adjetivo "bueno" (BANGOR; KORTUM; MILLER, 2009) sirven como parámetros interpretativos, y no como un límite universal de aprobación.

El SUS se complementará con indicadores como la correcta realización de tareas, errores, comprensión de qué es y qué no se analiza y percepción de privacidad. Al tratarse de adolescentes, el plan se someterá a evaluación ética institucional antes del reclutamiento, con el consentimiento de los responsables, el consentimiento de los adolescentes y la protección de

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

los registros; el marco debe confirmarse con el Comité de Ética de la Investigación, incluida la Resolución del CNS No. 510/2016 (BRASIL, 2016).

### 5.3 Gobernanza de datos, privacidad y trabajo futuro

La cuarta etapa es especificar una política de datos coherente con la arquitectura y los derechos de los niños, niñas y adolescentes. Incluso sin la transmisión del contenido de los mensajes, los metadatos de alerta como categoría, fuente y hora pueden ser datos personales cuando están relacionados con un menor, y el propósito, la necesidad y el mejor interés deben evaluarse para cada dato y etapa de tratamiento. (BRASIL, 2018, art. 14; BRASIL, 2025, arts. 7º e 16).

Ley No. 15.211/2025 determina que el controlador mapea y busca mitigar los riesgos del procesamiento de datos de niños y adolescentes y elaborar un informe de impacto, y el Decreto No. 12.880/2026 se ocupa de la evaluación de impacto en la seguridad y la salud de los niños. (BRASIL, 2025, art. 16; BRASIL, 2026, art. 47). El proyecto debe examinar la incidencia de estos requisitos en su diseño y concluir la evaluación de riesgos antes de cualquier piloto, como una medida de gobernanza de su propio trabajo. La arquitectura también debe analizarse a la luz de la prohibición de mecanismos de vigilancia masiva, genérica o indiscriminada. (BRASIL, 2025, art. 34, § 1º, e art. 37, parágrafo único), evaluando a qué contenidos llega el sistema, qué metadatos retiene y quién puede consultar alertas.

## 6 Consideraciones finales

La presente investigación tuvo como objetivo analizar los mecanismos tecnológicos de protección de niños, niñas y adolescentes en el entorno digital, considerando los riesgos asociados al ciberacoso y otras conductas nocivas, así como lo dispuesto en la Ley General de Protección de Datos Personales (LGPD) y el Estatuto Digital de la Niñez y la Adolescencia (ECA Digital). A partir del análisis realizado, fue posible comprender la importancia de conciliar seguridad, privacidad y supervisión parental en la protección de este público.

El análisis de los mecanismos de seguridad de las plataformas Discord, Instagram, TikTok y WhatsApp mostró que, aunque existen recursos para la moderación de contenido, la verificación de edad y el control parental, estas herramientas tienen limitaciones, especialmente en la identificación de situaciones de riesgo en conversaciones privadas. Así, se observó la necesidad de soluciones complementarias que ayuden a los responsables de la prevención de conductas nocivas, respetando la privacidad y autonomía de los niños, niñas y adolescentes.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

En este contexto, se desarrolló el prototipo de la aplicación Guarda Mirim, una propuesta tecnológica destinada a apoyar la supervisión parental. El trabajo dio como resultado la definición de requisitos, la elaboración de una arquitectura y la implementación de siete pantallas navegables, con funcionalidades destinadas a presentar alertas y pautas a los responsables. La propuesta busca permitir la identificación de situaciones de riesgo sin exponer el contenido original de las conversaciones, priorizando los principios de seguridad y protección de datos.

Sin embargo, es importante destacar que el prototipo utiliza datos simulados y aún no tiene implementados mecanismos de detección automatizados. Además, no se realizaron pruebas con familias ni evaluaciones prácticas de su efectividad. Por lo tanto, la hipótesis presentada se mantuvo en el ámbito conceptual, pero aún depende de la validación. Como continuación de la investigación, se recomienda investigar la viabilidad de la clasificación automatizada de contenidos, realizar pruebas de usabilidad con tutores y adolescentes y profundizar los mecanismos de privacidad y protección de datos.

Finalmente, se enfatiza que la tecnología no reemplaza el diálogo familiar y la educación digital, sino que puede actuar como una herramienta para apoyar la construcción de un entorno virtual más seguro.

## Referencias

AGÊNCIA BRASIL. ECA Digital começa a valer nesta terça; confira principais pontos. Brasília, 17 mar. 2026. Disponible en: https://agenciabrasil.ebc.com.br/politica/noticia/2026-03/eca-digital-comeca-valer-nesta-terca-confira-principais-pontos. Consultado el: 5 set. 2026.

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). Plataformas digitais acessadas por crianças e adolescentes precisam publicar relatório de transparência até 17/09. Brasília, DF, 11 ago. 2026. Disponible en: https://www.gov.br/anpd/pt-br/assuntos/noticias/plataformas-digitais-acessadas-por-criancas-e-adolescentes-precisam-publicar-relatorio-de-transparencia-ate-17-de-setembro. Consultado el: 18 out. 2026.

BANGOR, Aaron; KORTUM, Philip; MILLER, James. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability Studies, v. 4, n. 3, p. 114–123, maio 2009. Disponible en: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Bangor_May2009.pdf. Consultado el: 18 out. 2026.

BARDIN, Laurence. Análise de conteúdo. São Paulo: Edições 70, 2016.

BRASIL. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Diário Oficial da União, Brasília, DF, 24 maio 2016.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

BRASIL. Decreto nº 12.880, de 18 de março de 2026. Regulamenta a Lei nº 15.211, de 17 de setembro de 2025, e institui a Política Nacional de Promoção e Proteção dos Direitos da Criança e do Adolescente no Ambiente Digital. Diário Oficial da União, Brasília, DF, 18 mar. 2026, edição extra. Disponible en: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/decreto/d12880.htm. Consultado el: 18 out. 2026.

BRASIL. Lei nº 13.185, de 6 de novembro de 2015. Institui o Programa de Combate à Intimidação Sistemática (Bullying). Diário Oficial da União, Brasília, DF, 9 nov. 2015. Disponible en: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13185.htm. Consultado el: 18 out. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, Brasília, DF, 15 ago. 2018. Disponible en: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Consultado el: 19 out. 2026.

BRASIL. Lei nº 14.811, de 12 de janeiro de 2024. Institui medidas de proteção à criança e ao adolescente contra a violência nos estabelecimentos educacionais ou similares, prevê a Política Nacional de Prevenção e Combate ao Abuso e Exploração Sexual da Criança e do Adolescente e altera o Código Penal e outras leis. Diário Oficial da União, Brasília, DF, 15 jan. 2024. Disponible en: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/lei/l14811.htm. Consultado el: 19 out. 2026.

BRASIL. Lei nº 15.211, de 17 de setembro de 2025. Dispõe sobre a proteção de crianças e adolescentes em ambientes digitais (Estatuto Digital da Criança e do Adolescente). Diário Oficial da União, Brasília, DF, 17 set. 2025, edição extra. Disponible en: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/l15211.htm. Consultado el: 19 out. 2026.

BROOKE, John. SUS: a “quick and dirty” usability scale. In: JORDAN, Patrick W. et al. (ed.). Usability evaluation in industry. London: Taylor & Francis, 1996. p. 189–194.

CANALTECH. WhatsApp ganha novos controles para pais de adolescentes: veja o que muda. [S. l.], 30 set. 2026. Disponible en: https://canaltech.com.br/apps/whatsapp-ganha-novos-controles-para-pais-de-adolescentes-veja-o-que-muda/. Consultado el: 1 nov. 2026.

DISCORD. Discord Family Center. [S. l.], [2026]. Disponible en: https://discord.com/safety-family-center. Consultado el: 1 nov. 2026.

FANTTI, Bruna. Morte de adolescente de 13 anos leva polícia a grupo suspeito de induzir jovens ao suicídio pela internet. Jornal de Brasília, Brasília, DF, 4 ago. 2026. Reportagem da Folhapress. Disponible en: https://jornaldebrasilia.com.br/noticias/brasil/morte-de-adolescente-de-13-anos-leva-policia-a-grupo-suspeito-de-induzir-jovens-ao-suicidio-pela-internet/. Consultado el: 1 nov. 2026.

FEDERAL TRADE COMMISSION. Children’s Online Privacy Protection Rule (COPPA). Washington, DC, 2025. Disponible en:

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa. Consultado el: 18 out. 2026.

FUNDAÇÃO ABRINQ PELOS DIREITOS DA CRIANÇA E DO ADOLESCENTE. ECA Digital: o que muda para crianças e adolescentes. São Paulo, 6 jul. 2026. Disponible en: https://www.fadc.org.br/noticias/eca-digital-o-que-muda. Consultado el: 5 set. 2026.

GHOSH, Arup Kumar et al. A matter of control or safety? Examining parental use of technical monitoring apps on teens' mobile devices. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (CHI), 2018, Montreal. Proceedings [...]. New York: ACM, 2018b. DOI: 10.1145/3173574.3173768.

GHOSH, Arup Kumar et al. Safety vs. surveillance: what children have to say about mobile apps for parental control. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (CHI), 2018, Montreal. Proceedings [...]. New York: ACM, 2018a. DOI: 10.1145/3173574.3173698.

GIL, Antonio Carlos. Como elaborar projetos de pesquisa. 6. ed. São Paulo: Atlas, 2017.

HARD, Andrew et al. Federated learning for mobile keyboard prediction. arXiv, arXiv:1811.03604, 2018. Disponible en: https://arxiv.org/abs/1811.03604. Consultado el: 18 out. 2026.

HEVNER, Alan R. et al. Design science in information systems research. MIS Quarterly, v. 28, n. 1, p. 75-105, 2004. DOI: 10.2307/25148625.

LEITE, João Augusto et al. Toxic language detection in social media for Brazilian Portuguese: new dataset and multilingual analysis. In: CONFERENCE OF THE ASIA-PACIFIC CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (AACL-IJCNLP), 1., 2020, Suzhou. Proceedings [...]. [S. l.]: ACL, 2020. p. 914–924. Disponible en: https://arxiv.org/abs/2010.04543. Consultado el: 18 out. 2026.

LINGUISTICS, 58., 2020. Proceedings [...]. [S. l.]: ACL, 2020. p. 2158–2170. DOI: 10.18653/v1/2020.acl-main.195.

MCLINDON, Elizabeth et al. Timelines of psychological, physical and sexual intimate partner violence among a nationally representative sample of Australian women. Women's Health, v. 21, p. 1-12, 2025. DOI: 10.1177/17455057251329640.

MCMAHAN, Brendan et al. Communication-efficient learning of deep networks from decentralized data. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS (AISTATS), 20., 2017, Fort Lauderdale. Proceedings [...]. [S. l.]: PMLR, 2017. v. 54, p. 1273–1282.

META. Apresentamos as contas gerenciadas por pais no WhatsApp. [S. l.], 11 mar. 2026a. Disponible en: https://about.fb.com/br/news/2026/03/apresentamos-as-contas-gerenciadas-por-pais-no-whatsapp/. Consultado el: 1 nov. 2026.

META. Iniciativas para manter pais brasileiros informados sobre as contas de adolescente e supervisão parental no Instagram. [S. l.], 15 jun. 2026b (atualizado em 29 out. 2026).

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

Disponible en: https://about.fb.com/br/news/2026/06/iniciativas-para-manter-pais-brasileiros-informados-sobre-as-contas-de-adolescente-e-supervisao-parental-no-instagram/. Consultado el: 1 nov. 2026.

META. Supervisão parental simplificada em um só lugar. [S. l.], 12 maio 2026c. Disponible en: https://about.fb.com/br/news/2026/05/supervisao-parental-simplificada-em-um-so-lugar/. Consultado el: 18 out. 2026.

MIGALHAS. Governo processa Discord, pede R$ 500 mi e cobra proteção a menores. [S. l.], 26 ago. 2026. Disponible en: https://www.migalhas.com.br/quentes/463309/governo-processa-discord-pede-r-500-mi-e-cobra-protecao-a-menores. Consultado el: 5 set. 2026.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR (NIC.br) (ed.). Pesquisa sobre o uso da Internet por crianças e adolescentes no Brasil: TIC Kids Online Brasil 2025. 1. ed. São Paulo: Comitê Gestor da Internet no Brasil, 2026. ISBN 978-65-6165-007-6.

PEFFERS, Ken et al. A design science research methodology for information systems research. Journal of Management Information Systems, v. 24, n. 3, p. 45-77, 2007. DOI: 10.2753/MIS0742-1222240302.

PIAUÍ HOJE. MPF investiga segurança do Discord para proteger menores. [S. l.], 19 ago. 2026. Disponible en: https://piauihoje.com/index.php/noticias/brasil/mpf-investiga-seguranca-do-discord-para-proteger-menores-463658.html. Consultado el: 5 set. 2026.

SALAWU, Semiu; HE, Yulan; LUMSDEN, Joanna. Approaches to automated detection of cyberbullying: a survey. IEEE Transactions on Affective Computing, v. 11, n. 1, p. 3–24, 2020. DOI: 10.1109/TAFFC.2017.2761757.

SHOWMETECH. Governo federal proíbe o Discord para menores de 18 anos. [S. l.], 15 jun. 2026. Disponible en: https://www.showmetech.com.br/governo-federal-proibe-o-discord-para-menores-de-18-anos/. Consultado el: 5 set. 2026.

SOUZA, Fábio; NOGUEIRA, Rodrigo; LOTUFO, Roberto. BERTimbau: pretrained BERT models for Brazilian Portuguese. In: BRAZILIAN CONFERENCE ON INTELLIGENT SYSTEMS (BRACIS), 9., 2020, Rio Grande. Proceedings [...]. Cham: Springer, 2020. (Lecture Notes in Computer Science, v. 12319), p. 403–417.

SUN, Zhiqing et al. MobileBERT: a compact task-agnostic BERT for resource-limited devices. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL

TIKTOK. Nosso compromisso inequívoco com a segurança dos jovens no TikTok. [S. l.], 17 mar. 2026. Disponible en: https://newsroom.tiktok.com/nosso-compromisso-inequvoco-com-a-segurana-dos-jovens-no-tiktok?lang=pt-BR. Consultado el: 1 nov. 2026.

TIKTOK. Privacidade e segurança para usuários menores de 18 anos. [S. l.], [2026?]. Disponible en: https://support.tiktok.com/pt_BR/account-and-privacy/account-privacy-settings/privacy-and-safety-settings-for-users-under-age-18. Consultado el: 18 out. 2026.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016. Relativo à proteção das pessoas singulares no que diz respeito ao

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Ano VII, v.1 2026 | submissão: 15/11/2026 | aceito: 22/11/2026 | publicação: 27/11/2026**

tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Jornal Oficial da União Europeia, Luxemburgo, L 119, p. 1-88, 4 maio 2016. Disponible en: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Consultado el: 18 out. 2026.

VARGAS, Francielle et al. HateBR: a large expert annotated corpus of Brazilian Instagram comments for offensive language and hate speech detection. In: LANGUAGE RESOURCES AND EVALUATION CONFERENCE (LREC), 13., 2022, Marseille. Proceedings [...]. Paris: ELRA, 2022. p. 7174–7183. Disponible en: https://aclanthology.org/2022.lrec-1.777. Consultado el: 18 out. 2026.

WHATSAPP. Como manter a segurança no WhatsApp. [S. l.], [2026?]. Disponible en: https://faq.whatsapp.com/1313491802751163/?locale=pt_BR. Consultado el: 18 out. 2026.

WHITTLE, Helen et al. A review of online grooming: characteristics and concerns. Aggression and Violent Behavior, v. 18, n. 1, p. 62-70, 2013.

WIENER, Cassandra. Seeing what is ‘invisible in plain sight’: policing coercive control. The Howard Journal of Crime and Justice, v. 56, n. 4, p. 500-515, 2017. DOI: 10.1111/hojo.12227.

WISNIEWSKI, Pamela et al. Parental control vs. teen self-regulation: is there a middle ground for mobile online safety? In: ACM CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK AND SOCIAL COMPUTING (CSCW), 2017, Portland. Proceedings [...]. New York: ACM, 2017. p. 51–69. DOI: 10.1145/2998181.2998352.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.