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

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

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

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

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 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,

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

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.

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

1 Introdução

No cenário contemporâneo, o avanço tecnológico tornou o acesso à internet por crianças e adolescentes cada vez mais comum, modificando a forma como esse público se comunica e interage com o mundo. Segundo a pesquisa TIC Kids Online Brasil 2025, realizada pelo Núcleo de Informação e Coordenação do Ponto BR (NIC.br), 92% das crianças e adolescentes brasileiros entre 9 e 17 anos eram usuários da internet, evidenciando a presença das tecnologias

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

digitais em seu cotidiano (NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR, 2026).

Apesar dos benefícios proporcionados pela conectividade, a crescente presença desse público no ambiente digital também amplia sua exposição a riscos como cyberbullying, contato com desconhecidos, manipulação, coerção e conteúdos inadequados. Diante desse cenário, a proteção de crianças e adolescentes exige acompanhamento familiar e mecanismos tecnológicos capazes de auxiliar na identificação e prevenção de comportamentos prejudiciais.

Nesse contexto, a proteção infantojuvenil deve considerar as disposições legais relacionadas à segurança e à privacidade. A Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709/2018, estabelece que o tratamento de dados de crianças e adolescentes deve observar seu melhor interesse. Complementarmente, o Estatuto Digital da Criança e do Adolescente (ECA Digital), instituído pela Lei nº 15.211/2025, estabelece medidas de proteção e responsabilidades para os agentes envolvidos na oferta de serviços tecnológicos.

Embora plataformas como Discord, Instagram, TikTok e WhatsApp disponibilizem mecanismos de segurança e supervisão parental, essas ferramentas apresentam limitações, especialmente na identificação de riscos em conversas privadas. Considerando essas necessidades, a pesquisa propõe o aplicativo Guarda Mirim, uma ferramenta complementar de proteção com mecanismos previstos de detecção e alerta, buscando preservar a privacidade dos menores sem expor o conteúdo original das conversas. Nesta etapa, a proposta é apresentada por meio de um protótipo navegável com dados simulados.

Diante desse cenário, estabelece-se o seguinte problema de pesquisa: de que forma mecanismos tecnológicos de detecção e prevenção podem contribuir para a proteção de crianças e adolescentes contra o cyberbullying e outros comportamentos de risco no ambiente digital, considerando a LGPD, o ECA Digital e a participação dos responsáveis na supervisão?

O objetivo geral desta pesquisa consiste em analisar mecanismos tecnológicos de proteção de crianças e adolescentes em ambientes digitais, com ênfase na prevenção do cyberbullying e de outros comportamentos de risco, considerando a legislação brasileira e propondo o aplicativo Guarda Mirim como ferramenta complementar de supervisão parental.

Para alcançar esse objetivo, foram estabelecidos os seguintes objetivos específicos: identificar os principais riscos e formas de violência digital contra crianças e adolescentes; analisar os mecanismos de proteção das plataformas Discord, Instagram, TikTok e WhatsApp, considerando as exigências da LGPD e do ECA Digital; e apresentar os requisitos, a arquitetura

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

proposta e o protótipo de interface do Guarda Mirim, respeitando os princípios de segurança e privacidade.

Parte-se da hipótese de que mecanismos tecnológicos de detecção e prevenção, desenvolvidos conforme os princípios de segurança, privacidade e melhor interesse de crianças e adolescentes, podem contribuir para a identificação antecipada de situações de risco e fortalecer a atuação dos responsáveis, sem substituir o acompanhamento familiar.

Quanto aos procedimentos metodológicos, a pesquisa caracteriza-se como aplicada, qualitativa, exploratória e descritiva, fundamentada em pesquisas bibliográficas e documentais e na análise comparativa dos mecanismos de proteção de quatro plataformas digitais. O desenvolvimento do protótipo foi orientado pela abordagem design science research, contemplando a definição de requisitos, a modelagem e a implementação das interfaces.

Como resultado, foram identificadas limitações nos mecanismos de proteção das plataformas e desenvolvido um protótipo navegável do Guarda Mirim, composto por sete telas. Entretanto, a detecção automatizada ainda não foi implementada, sendo necessárias avaliações futuras para verificar sua eficácia.

Por fim, o artigo está organizado em cinco seções, além desta introdução: referencial teórico, metodologia, desenvolvimento do protótipo Guarda Mirim, resultados e discussão, e considerações finais.

2 Marco teórico

2.1 Riscos e violências digitais contra crianças e adolescentes

Em 2025, 92% das pessoas de 9 a 17 anos no Brasil eram usuárias de internet, cerca de 24 milhões de crianças e adolescentes (NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR, 2026). Para esse público, a internet é espaço de estudo, lazer e socialização, mas também de exposição a riscos que dependem não apenas do conteúdo acessado, mas de quem entra em contato com o menor e de como as relações se desenvolvem online.

O cyberbullying é uma forma de violência digital. Nos termos da Lei nº 13.185/2015, “há intimidação sistemática na rede mundial de computadores (cyberbullying), quando se usarem os instrumentos que lhe são próprios para depreciar, incitar a violência, adulterar fotos e dados pessoais com o intuito de criar meios de constrangimento psicossocial” (BRASIL, 2015, art. 2º, parágrafo único). A Lei nº 14.811/2024 passou a tratar a intimidação sistemática, inclusive a virtual, como crime (BRASIL, 2024). Em comparação com o bullying presencial, o

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

cyberbullying pode alcançar públicos ampliados, manter registros acessíveis e ocorrer sob anonimato, características relevantes para sua detecção e enfrentamento (SALAWU; HE; LUMSDEN, 2020).

Para fins desta pesquisa, os riscos foram organizados em quatro categorias, que também orientam o protótipo GuardaMirim:

- Cyberbullying: agressões repetidas, humilhação e exclusão de um menor por meio de mensagens, publicações ou grupos.
- Linguagem ofensiva: xingamentos, discurso de ódio e ameaças, que podem ocorrer de forma isolada, sem a repetição típica do cyberbullying, mas com potencial de dano.
- Imagem inadequada: recebimento ou compartilhamento de imagens impróprias para a idade, como nudez e violência, inclusive imagens íntimas do próprio menor.
- Contato desconhecido: aproximação de adultos ou desconhecidos, que pode evoluir para aliciamento.

A manipulação e a coerção não constituem necessariamente categorias isoladas, mas estratégias transversais que podem estar presentes em diferentes manifestações de violência. Estudos sobre o aliciamento online de crianças e adolescentes e sobre o controle coercitivo apontam que essas práticas podem se desenvolver progressivamente, por meio da construção de confiança, do controle emocional e do isolamento social da vítima, estabelecendo condições que favorecem o domínio exercido pelo agressor (WHITTLE et al., 2013; WIENER, 2017; MCLINDON et al., 2025). Esse padrão aparece, por exemplo, em investigação policial noticiada em 2026, na qual os suspeitos, após ganharem a confiança de crianças e adolescentes, submetiam-nos a coação, isolamento e violência psicológica (FANTTI, 2026). As categorias também não são excludentes, pois uma mesma situação pode combinar, por exemplo, contato desconhecido, imagem inadequada e coerção.

2.2 Marco legal da proteção de crianças e adolescentes no ambiente digital

No Brasil, a Lei Geral de Proteção de Dados Pessoais (LGPD) dedica o art. 14 ao tratamento de dados de crianças e adolescentes, que deve ser realizado em seu melhor interesse (BRASIL, 2018). No caso de crianças, exige-se consentimento específico e em destaque de ao menos um dos pais ou do responsável legal, e as informações sobre o tratamento devem ser apresentadas em linguagem simples e acessível (BRASIL, 2018, art. 14, §§ 1º e 6º). Aplicam-se ainda os princípios do art. 6º, em especial finalidade, necessidade, transparência, segurança

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

e prevenção. Para soluções tecnológicas, isso significa coletar apenas os dados estritamente necessários e explicar com clareza o que é tratado.

O Estatuto Digital da Criança e do Adolescente (ECA Digital), instituído pela Lei nº 15.211/2025, entrou em vigor em 17 de março de 2026 (AGÊNCIA BRASIL, 2026). Aplica-se a produtos e serviços de tecnologia da informação direcionados a crianças e adolescentes ou de acesso provável por eles, o que abrange redes sociais, aplicativos de mensagens e plataformas de jogos (BRASIL, 2025; FUNDAÇÃO ABRINQ PELOS DIREITOS DA CRIANÇA E DO ADOLESCENTE, 2026). Entre suas principais exigências estão:

- verificação de idade por mecanismos confiáveis, sendo insuficiente a mera autodeclaração;
- supervisão parental, com ferramentas acessíveis para que pais e responsáveis acompanhem e configurem o uso, incluindo a vinculação das contas de adolescentes de até 16 anos à de um responsável;
- transparência, com relatórios periódicos das plataformas (art. 31), cujo primeiro prazo de publicação foi fixado pela Agência Nacional de Proteção de Dados em 17 de setembro de 2026 (AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS, 2026).

A lei estabelece também que o cumprimento dessas obrigações não pode resultar em mecanismos de vigilância massiva, genérica ou indiscriminada (BRASIL, 2025, art. 34, § 1º). Essa previsão é central para este estudo, porque delimita o que uma solução de proteção pode fazer sem violar a privacidade do menor.

No cenário internacional, a Children's Online Privacy Protection Rule (COPPA), dos Estados Unidos, exige consentimento parental verificável para a coleta de dados de menores de 13 anos (FEDERAL TRADE COMMISSION, 2025), e o art. 8º do Regulamento Geral de Proteção de Dados da União Europeia (GDPR) fixa em 16 anos a idade do consentimento, podendo os Estados-membros reduzi-la até 13 (UNIÃO EUROPEIA, 2016). Esses marcos concentram-se no consentimento e no tratamento de dados, enquanto o ECA Digital trata também do desenho dos serviços, da verificação etária e da supervisão parental.

2.3 Mecanismos de proteção em ambientes digitais

A partir do marco legal, identificam-se cinco mecanismos de proteção, que correspondem aos critérios do Quadro 1.

Verificação de idade: busca estabelecer a faixa etária do usuário para oferecer uma experiência adequada e restringir conteúdos impróprios. Diferencia-se da verificação 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

identidade, pois identifica se o usuário é menor, e não quem ele é. A autodeclaração é considerada insuficiente pelo ECA Digital (BRASIL, 2025).

Moderação de conteúdo: compreende a detecção automatizada, a revisão humana e os canais de denúncia usados para identificar e remover condutas nocivas. A literatura sobre detecção automatizada de cyberbullying aponta desafios de interpretação contextual e classificação (SALAWU; HE; LUMSDEN, 2020); esses sistemas podem errar e, em serviços com criptografia de ponta a ponta, a análise do conteúdo pela plataforma é tecnicamente restrita; por isso, a moderação isolada não basta para proteger menores em conversas privadas.

Controle parental: consiste em ferramentas que permitem aos responsáveis vincular contas, estabelecer limites de tempo de uso e de contatos, além de acompanhar determinados aspectos da utilização dos serviços. O ECA Digital estabelece a necessidade de disponibilização e acessibilidade dessas ferramentas (BRASIL, 2025). Entretanto, estudos sobre controle parental identificam tensões entre proteção, autonomia e privacidade dos adolescentes (WISNIEWSKI et al., 2017; GHOSH et al., 2018a).

Privacidade do conteúdo do menor: refere-se ao grau de acesso que a plataforma, os responsáveis ou terceiros possuem ao conteúdo das interações dos menores. Esse mecanismo está relacionado aos princípios do melhor interesse e da necessidade (BRASIL, 2018), bem como à vedação à vigilância massiva prevista no ECA Digital (BRASIL, 2025, art. 34, § 1º). Portanto, a supervisão deve buscar o menor nível de exposição de dados possível, sem comprometer a proteção necessária.

Resposta a incidentes e transparência: abrange a publicação de relatórios, o atendimento às denúncias e a submissão das plataformas à fiscalização e às sanções previstas em lei. Esses procedimentos permitem que a sociedade e o poder público avaliem, com base em documentos públicos, se os mecanismos de proteção são efetivos na prática (BRASIL, 2025, art. 31).

2.4 Síntese e lacuna

A literatura científica e a legislação examinadas apontam duas tendências (WISNIEWSKI et al., 2017; GHOSH et al., 2018a; BRASIL, 2025). A primeira é a responsabilização das plataformas por verificar a idade, moderar conteúdos e oferecer supervisão parental. A segunda é a preocupação com a privacidade do menor, que limita o acesso irrestrito ao conteúdo das conversas. Os mecanismos nativos, no entanto, operam dentro de cada serviço, o que obriga o responsável a acompanhar vários painéis, e nem sempre

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

conseguem alertá-lo a tempo sobre situações como cyberbullying, aliciamento e coerção, que se manifestam em conversas privadas.

Essa lacuna indica a necessidade de soluções complementares que identifiquem sinais de risco e avisem o responsável sem lhe entregar o conteúdo das conversas, em consonância com os princípios de minimização e de melhor interesse. É nesse espaço que se insere a proposta do GuardaMirim, detalhada na seção 4, e cuja adequação ao marco aqui apresentado é discutida adiante.

3 Material e método

3.1 Caracterização da pesquisa

Esta pesquisa é de natureza aplicada, com abordagem qualitativa e caráter exploratório e descritivo (GIL, 2017). Ela foi conduzida em duas frentes complementares. A primeira é uma análise documental comparativa dos mecanismos de segurança e moderação adotados por plataformas digitais para proteger crianças e adolescentes. A segunda é a construção de um artefato tecnológico, o protótipo GuardaMirim, orientada pela design science research, abordagem na qual o desenvolvimento de um artefato é parte do método de investigação (HEVNER et al., 2004; PEFFERS et al., 2007).

A opção pela análise documental justifica-se pelo caráter recente do objeto: a Lei nº 15.211/2025 (ECA Digital) entrou em vigor em 17 de março de 2026 (AGÊNCIA BRASIL, 2026), e as plataformas ainda ajustam seus recursos e documentos. Como os materiais diferem em escopo e nível de detalhamento, a comparação não pressupõe equivalência direta entre eles. A pesquisa não incluiu entrevistas nem coleta direta de dados junto às empresas, restringindo-se à análise de documentos públicos e ao desenvolvimento do protótipo.

3.2 Fontes de dados e critérios de seleção

Foram consultadas três categorias de fontes: (i) legislação e documentos oficiais, incluindo a Lei nº 15.211/2025 (BRASIL, 2025), a Lei nº 13.709/2018 (BRASIL, 2018), a regra norte-americana COPPA (FEDERAL TRADE COMMISSION, 2025), o Regulamento Geral de Proteção de Dados da União Europeia (UNIÃO EUROPEIA, 2016) e materiais da Agência Nacional de Proteção de Dados (ANPD); (ii) documentos institucionais de cada plataforma, selecionado entre materiais de segurança, ajuda, privacidade, supervisão parental e transparência do Discord, Instagram/Meta, TikTok e WhatsApp (DISCORD, [2026?]; META,

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

2026a, 2026b, 2026c; TIKTOK, [2026?], 2026; WHATSAPP, [2026?]); e (iii) materiais institucionais, relatórios de pesquisa e reportagens sobre fiscalização, litígios, proteção infantojuvenil e mudanças de recursos das plataformas.

As fontes primárias foram consultadas nos portais oficiais do Planalto, da ANPD, da Federal Trade Commission e do EUR-Lex, bem como nos centros de segurança e ajuda das plataformas. Para localizar estudos e notícias, utilizaram-se o Google Search, o Google Acadêmico, a SciELO e o Portal de Periódicos da CAPES, com termos como “ECA Digital”, “verificação de idade”, “controle parental” e “cyberbullying”, combinados por operadores AND e OR; a análise fundamentou-se nos documentos recuperados e verificados.

O corpus final foi composto por 20 documentos, contabilizados individualmente: quatro marcos normativos (ECA Digital, LGPD, COPPA e GDPR), sete materiais institucionais das plataformas (um do Discord, dois da Meta/Instagram, dois do WhatsApp e dois do TikTok), uma página oficial da ANPD sobre o ECA Digital, dois materiais institucionais/de pesquisa (Fundação Abrinq e NIC.br/Cetic.br) e seis reportagens (Agência Brasil, Migalhas, Piauí Hoje, Showmetech, Canaltech e Jornal de Brasília/Folhapress). As referências teóricas e metodológicas utilizadas apenas para fundamentar a pesquisa não foram incluídas nessa contagem.

Adotaram-se como critérios de inclusão: (a) publicação ou atualização entre 2025 e 2026 (para páginas sem data explícita, considerou-se a versão disponível na coleta, com registro da data de acesso), exceto marcos normativos anteriores (LGPD, COPPA e GDPR); (b) relação direta com ao menos um dos critérios da seção 3.4; e (c) identificação da autoria ou do órgão responsável. Priorizaram-se fontes primárias; as reportagens foram usadas de forma complementar, como relatos jornalísticos e não como fatos auditados. A coleta principal ocorreu em setembro de 2026, e a conferência final dos links foi feita até 1º de novembro de 2026.

3.3 Seleção das plataformas

Foram analisadas quatro plataformas: Discord, Instagram, TikTok e WhatsApp. O recorte baseou-se em (i) sua relevância no debate público e regulatório brasileiro em 2026, incluindo a ação civil pública da Advocacia-Geral da União contra o Discord (MIGALHAS, 2026), a investigação do Ministério Público Federal sobre a segurança do Discord (PIAUÍ HOJE, 2026) e a alteração da classificação indicativa (SHOWMETECH, 2026); e (ii) seu enquadramento na lei, que abrange redes sociais, aplicativos de mensagens e plataformas 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

jogos (FUNDAÇÃO ABRINQ PELOS DIREITOS DA CRIANÇA E DO ADOLESCENTE, 2026). Trata-se de amostragem intencional, sem pretensão de representatividade, e as diferenças entre os tipos de serviço limitam a comparabilidade direta.

3.4 Critérios de análise

A comparação foi estruturada em cinco critérios, derivados do referencial teórico (Quadro 1).

Quadro 1 – Critérios de análise e pergunta orientadora

Critério	Pergunta orientadora
Verificação de idade	A plataforma adota mecanismo além da autodeclaração?
Moderação de conteúdo	Há moderação automatizada, humana e canal de denúncia? Quais limites aparecem nas fontes?
Controle parental	Existe supervisão parental nativa? O que o responsável consegue ver e configurar?
Privacidade do conteúdo do menor	O responsável ou a plataforma acessam o conteúdo das interações?
Resposta a incidentes e transparência	Há relatórios, sanções ou ações regulatórias públicas?

Fonte: elaborado pelos autores (2026).

3.5 Procedimentos

O procedimento seguiu quatro etapas: (1) levantamento bibliográfico e normativo e definição dos critérios; (2) coleta e organização das informações em uma matriz plataforma × critério; (3) comparação dos mecanismos identificados com as exigências do ECA Digital e da LGPD; e (4) interpretação por análise de conteúdo categorial (BARDIN, 2016), com os cinco critérios do Quadro 1 como categorias. Cada célula foi classificada como mecanismo identificado em fonte oficial, identificado apenas em fonte jornalística, ou não identificado nas fontes consultadas, o que indica ausência de informação pública, e não ausência do recurso.

3.6 Aspectos éticos e legais

A pesquisa utilizou exclusivamente documentos de acesso público e não envolveu seres humanos. Por isso, não foi necessária aprovação de comitê de ética, nos termos do art. 1º, parágrafo único, inciso II, da Resolução CNS nº 510/2016 (BRASIL, 2016). Não foram coletados dados pessoais de crianças ou adolescentes: alertas, nomes e métricas do protótipo são fictícios. O desenho do artefato considerou o melhor interesse da criança e a minimização

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

de dados (BRASIL, 2018, art. 14) e a vedação à vigilância massiva (BRASIL, 2025, art. 34, § 1º). Qualquer estudo futuro com participantes exigirá avaliação ética prévia ao recrutamento.

4 Desenvolvimento do protótipo GuardaMirim

4.1 Visão geral e escopo

O GuardaMirim é um aplicativo de apoio à proteção infantojuvenil, projetado para detectar sinais de cyberbullying e outros riscos digitais e alertar o responsável sem expor o conteúdo das conversas do menor. O desenvolvimento seguiu três etapas iterativas, típicas da design science research: identificação do problema, projeto e desenvolvimento, e avaliação, que resumem as seis atividades do modelo original (PEFFERS et al., 2007).

O escopo do artefato limita-se, nesta etapa, ao projeto de interface e ao fluxo de navegação. O motor de classificação de conteúdo não foi implementado, e todos os dados exibidos são simulados. A arquitetura completa, descrita na seção 4.3, é uma proposta para etapas futuras.

4.2 Requisitos

Os requisitos foram derivados dos critérios de análise (Quadro 1) e das restrições legais discutidas no referencial teórico. O Quadro 2 apresenta os principais.

Quadro 2 – Requisitos do GuardaMirim

Código	Tipo	Descrição
RF01	Funcional	Permitir a seleção entre perfil do menor e perfil do responsável
RF02	Funcional	Exigir autenticação (PIN de 4 dígitos, adotado apenas como barreira mínima no protótipo) para acesso ao perfil do responsável
RF03	Funcional	Exibir status de proteção do dispositivo e ocorrências por categoria nos últimos 7 dias
RF04	Funcional	Listar alertas recentes com severidade (baixa, média ou alta), aplicativo de origem e horário
RF05	Funcional	Detalhar o alerta com metadados e resumo da classificação
RF06	Funcional	Oferecer ao responsável as ações: marcar como resolvido, consultar orientações de conversa e pausar o aplicativo de origem por 1 hora (requisito conceitual, sujeito às permissões do sistema operacional)
RF07	Funcional	Informar ao menor, em painel próprio, que a proteção está ativa e o que é analisado
RF08	Funcional	Permitir ao responsável ajustar a sensibilidade da detecção e ativar ou desativar categorias de alerta

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

Código	Tipo	Descrição
RF09	Funcional	Notificar o responsável sobre novos alertas, inclusive na tela de bloqueio, sem exibir o conteúdo bruto
RF10	Funcional	Permitir que o responsável registre concordância ou discordância com a classificação do alerta (retroalimentação)
RNF01	Não funcional	Privacidade por padrão: nenhuma tela pode exibir o conteúdo bruto de mensagens ou imagens
RNF02	Não funcional	Comunicar severidade por cor e texto simultaneamente, com tons não alarmistas
RNF03	Não funcional	Acessibilidade: foco visível, alvos de toque amplos e respeito à preferência por movimento reduzido

Fonte: elaborado pelos autores (2026).

As categorias de risco do protótipo (cyberbullying, linguagem ofensiva, imagem inadequada e contato desconhecido) refletem a taxonomia de riscos apresentada na seção 2.1.

4.3 Arquitetura proposta

A arquitetura foi concebida em camadas, de modo que o conteúdo das interações permaneça no dispositivo do menor (Figura 1 e Quadro 3). Apenas a classificação do evento (categoria, severidade, aplicativo de origem, contexto e horário) é encaminhada ao responsável.

Quadro 3 – Camadas da arquitetura e situação de implementação

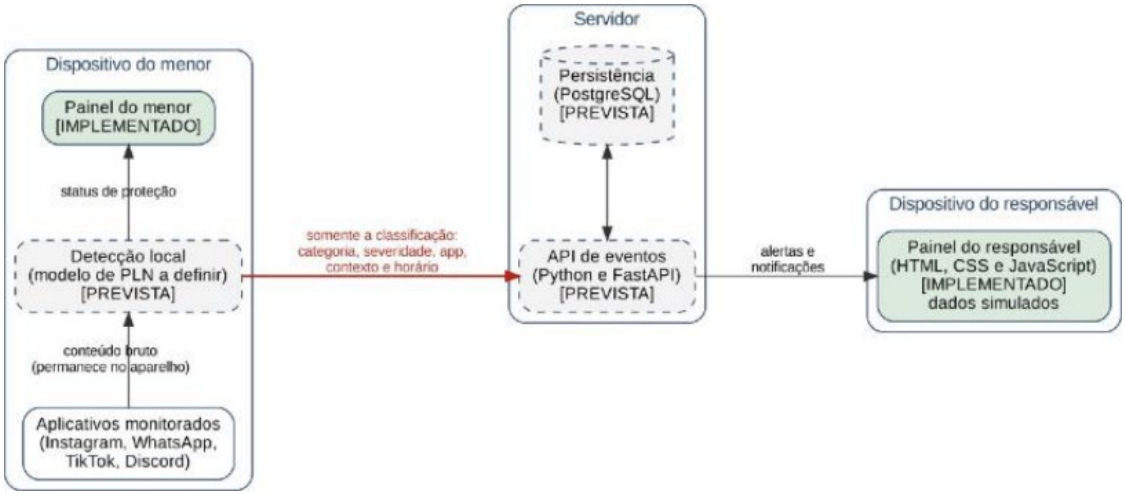
Camada	Função	Tecnologia	Situação
Interface (perfis menor e responsável)	Telas e navegação	HTML, CSS e JavaScript	Implementada (dados simulados)
Detecção local	Classificar o conteúdo no aparelho do menor	Modelo de processamento de linguagem natural (PLN) a definir	Prevista
API de eventos	Receber e entregar classificações de eventos	Python e FastAPI	Prevista
Persistência	Armazenar eventos e preferências	PostgreSQL	Prevista

Fonte: elaborado pelos autores (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

Figura 1 – Arquitetura proposta do GuardaMirim



Fonte: elaborado pelos autores (2026). Os blocos tracejados indicam componentes previstos, ainda não implementados.

Essa organização busca atender ao ECA Digital, que veda mecanismos de vigilância massiva, genérica ou indiscriminada (BRASIL, 2025, art. 34, § 1º). O que circula entre as camadas é a classificação do evento, nunca o texto ou a imagem que o originou. As políticas de retenção e descarte dos metadados de eventos ainda precisam ser definidas, com prazos, finalidade, controle de acesso, exclusão e registro de auditoria.

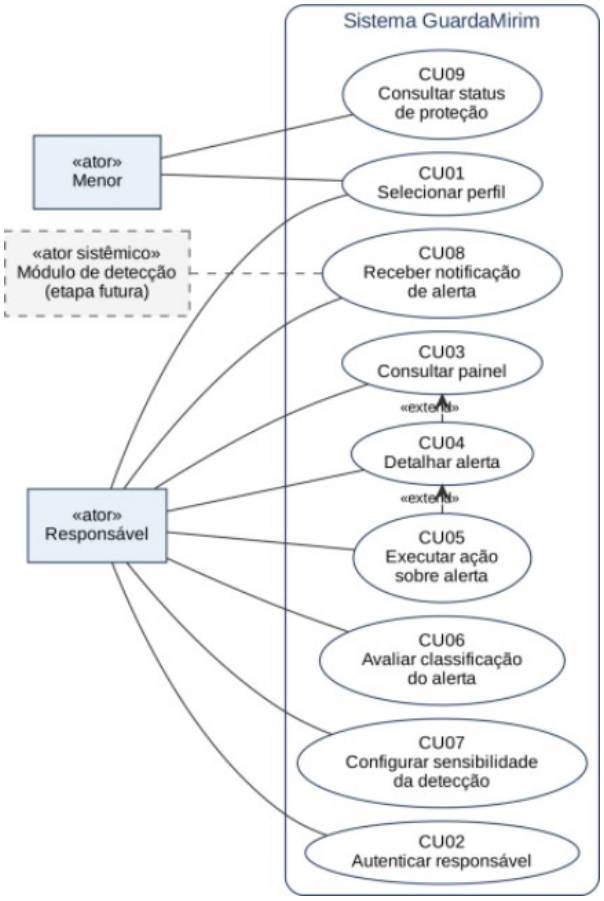
4.4 Modelagem

A modelagem identificou dois atores humanos, o menor e o responsável, e o módulo de detecção como ator sistêmico futuro. Foram levantados nove casos de uso (CU01 a CU09), apresentados no diagrama da 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 do GuardaMirim



Fonte: elaborado pelos autores (2026).

4.5 Implementação e interfaces

O protótipo foi implementado em HTML, CSS e JavaScript, como fluxo navegável de sete telas. O Quadro 4 relaciona cada tela à sua função e à decisão de privacidade correspondente, e a Figura 3 reúne as capturas. O fluxo do responsável percorre os painéis (a) a (f): seleção de perfil, autenticação, painel, detalhes do alerta, configuração da sensibilidade e notificação na tela de bloqueio; o fluxo do menor corresponde ao painel (g). O protótipo está disponível mediante solicitação aos autores.

Quadro 4 – Telas do protótipo e decisões de privacidade

Tela	Função	Decisão de privacidade
Seleção de perfil (Fig. 3a)	Separar os acessos	Nota informando que a análise é local
Autenticação (Fig. 3b)	Proteger o painel do responsável	Barreira mínima antes de dados sensíveis

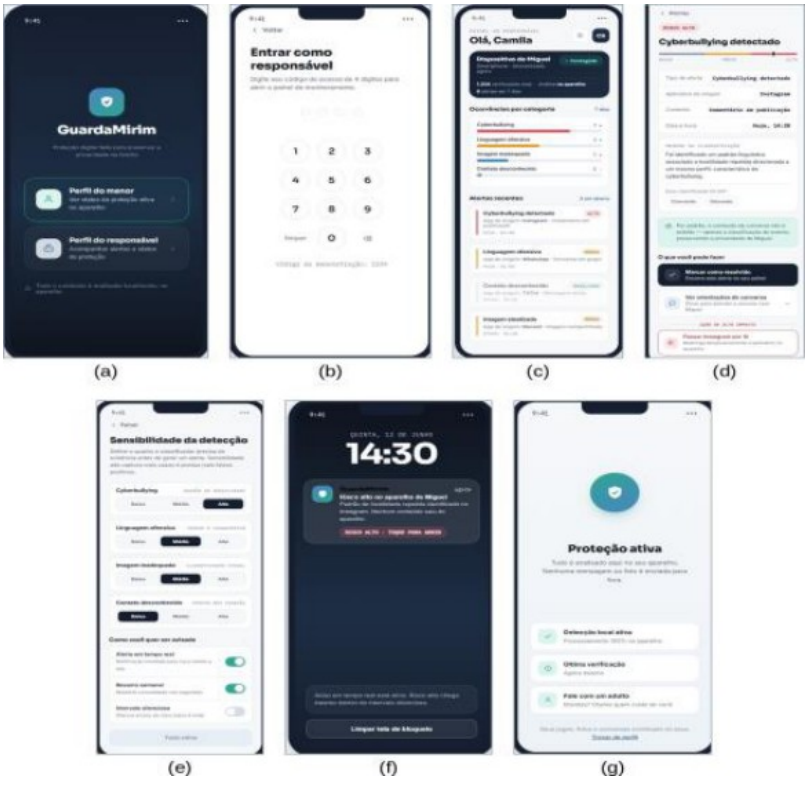
Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

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

Tela	Função	Decisão de privacidade
Painel do responsável (Fig. 3c)	Status, gráfico de categorias e feed de alertas	Exibe apenas metadados
Detalhes do alerta (Fig. 3d)	Metadados, resumo e ações	Aviso explícito de que o conteúdo não é exibido; ações voltadas à mediação familiar
Sensibilidade da detecção (Fig. 3e)	Ajustar o rigor da detecção e as categorias monitoradas	Controle do responsável sobre o nível de monitoramento, sem acesso ao conteúdo
Notificação (Fig. 3f)	Avisar o responsável sobre novo alerta na tela de bloqueio	Mensagem com categoria e severidade, sem trecho da conversa
Painel do menor (Fig. 3g)	Confirmar proteção ativa	Transparência sobre o que é analisado

Fonte: elaborado pelos autores (2026).

Figura 3 – Telas do protótipo GuardaMirim (a a g)



Fonte: elaborado pelos autores (2026).

4.6 Avaliação interna de consistência da interface

A avaliação desta etapa foi interna: a equipe revisou a coerência entre o discurso de privacidade comunicado ao usuário e o que cada tela de fato exibe. Dessa revisão resultou, por exemplo, a inclusão do aviso de privacidade na tela de detalhes do alerta. O protótipo ainda não

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

foi submetido a testes de usabilidade nem à validação com famílias, e não houve medição de desempenho de classificação.

5 Resultados e discussão

5.1 Comparação dos mecanismos de proteção nas plataformas

O Quadro 5 sintetiza a matriz plataforma × critério construída conforme a seção 3.5. Cada célula indica se o mecanismo foi identificado em fonte oficial (O), apenas em fonte jornalística (J) ou se não foi identificado nas fontes consultadas (–). Como previsto no método, a terceira situação indica ausência de informação nos documentos analisados, e não ausência do recurso na plataforma.

Quadro 5 – Matriz plataforma × critério: mecanismos identificados nas fontes analisadas

Critério	Discord	Instagram	TikTok	WhatsApp
Verificação de idade	J – cadastro com data de nascimento; a União alega que é insuficiente (MIGALHAS, 2026)	– Não identificado na fonte analisada (META, 2026b)	O – idade mínima de 13 anos; formulário de data de nascimento e modelos de detecção de contas de menores de 13 anos (TIKTOK, 2026)	– Não identificado; a fonte não descreve método (META, 2026a)
Moderação de conteúdo e denúncia	O – ferramentas de denúncia e bloqueio (DISCORD, [2026?]). J – a ação pede moderação em português e detecção de transmissões de risco (MIGALHAS, 2026)	O – canal de denúncia de conteúdos ilegais no Brasil, desde mar. 2026 (META, 2026b)	O – tecnologia e revisão humana; denúncia por qualquer pessoa; casos de exploração sexual infantil são reportados (TIKTOK, 2026)	O – denúncia e bloqueio de contatos e mensagens; aviso de contexto em mensagens de desconhecidos; banimento por violação dos Termos de Serviço, que proíbem conteúdo de abuso, intimidação ou ameaça (WHATSAPP, [2026?])
Controle parental	O – Family Center: vínculo de contas; o responsável vê amigos, servidores e contatos recentes (DISCORD, [2026?])	O – supervisão ativável sem autorização do adolescente; alertas de buscas sobre suicídio e automutilação (META, 2026b); gestão centralizada na Central da Família, com convite único para Instagram, Facebook, Messenger e Meta Horizon (META, 2026c)	O – Sincronização Familiar: limite de tempo, restrição de conteúdo e gestão de mensagens (TIKTOK, 2026)	O – contas gerenciadas por pais, para pré-adolescentes (META, 2026a). J – novos controles para adolescentes, anunciados em 30 set. 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

Critério	Discord	Instagram	TikTok	WhatsApp
Privacidade do conteúdo do menor	O – o responsável não acessa o conteúdo das mensagens privadas (DISCORD, [2026?])	– Não identificado; a fonte cita apenas tópicos de conversas com a Meta AI (META, 2026b)	O – contas de 13 a 17 anos privadas por padrão, com publicações visíveis apenas a seguidores aprovados; mensagens diretas indisponíveis para 13 a 15 anos (TIKTOK, [2026?]; TIKTOK, 2026). As fontes não esclarecem se os pais veem as mensagens	O – conversas com criptografia de ponta a ponta (META, 2026a). J – os pais não veem o conteúdo das conversas (CANALTECH , 2026)
Resposta a incidentes e transparência	J – ação civil pública da AGU, com pedido de indenização de ao menos R$ 500 milhões (MIGALHAS, 2026)	– Não identificado; a fonte não menciona relatórios de transparência (META, 2026b)	O – relatórios trimestrais no Centro de Transparência; acesso de pesquisadores a dados públicos de menores (TIKTOK, 2026)	– Não identificado nas fontes analisadas

Legenda: O = fonte oficial; J = apenas fonte jornalística; – = não identificado nas fontes consultadas. Fonte: elaborado pelos autores (2026), com base nos documentos citados em cada célula.

Observou-se que o controle parental e os canais de denúncia foram os critérios com mecanismo descrito em fonte oficial nas quatro plataformas. Em contraste, a verificação de idade foi descrita em fonte oficial apenas no TikTok. No caso do Discord, a única referência encontrada vem de fonte jornalística, que relata a alegação da União de que a declaração da data de nascimento é insuficiente, e, para Instagram e WhatsApp, as fontes analisadas não detalham o método. Quanto aos relatórios de transparência, apenas o TikTok os descreve em fonte oficial; sobre o Discord, a fonte encontrada trata de ação regulatória, e não de relatório.

Quanto à privacidade, Discord e WhatsApp informam que o responsável não acessa o conteúdo das conversas, e, no WhatsApp, a criptografia de ponta a ponta impede o acesso até mesmo da empresa, segundo a Meta (META, 2026a). Essa convergência é compatível com a vedação à vigilância massiva (BRASIL, 2025, art. 34, § 1º) e confirma a tensão entre proteção e privacidade da seção 2.3: os controles nativos entregam ao responsável configurações e informações de uso, mas não o conteúdo. Os alertas identificados referem-se a eventos específicos, como buscas por suicídio e automutilação no Instagram (META, 2026b) e entrada em grupos no WhatsApp (CANALTECH, 2026). Não foram identificadas, nas fontes consultadas, descrições públicas de alertas ao responsável sobre sinais de cyberbullying ou aliciamento em conversas privadas, lacuna que motiva o GuardaMirim.

Cabe ressaltar os limites da comparação. A matriz depende dos documentos disponíveis até 1º de novembro de 2026, e as plataformas alteram seus recursos com frequência; a nota da Meta sobre o Instagram, por exemplo, foi atualizada em 29 de outubro de 2026. As informações

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

sobre falhas do Discord decorrem de alegações da União reportadas pela imprensa, e não de auditoria.

5.2 Limitações do protótipo e etapas futuras de validação

Esta subseção apresenta as limitações do protótipo e as etapas futuras do GuardaMirim. Os resultados esperados aqui descritos não foram obtidos nesta pesquisa: são objetivos de desenvolvimento e avaliação. O protótipo demonstra a interface e o fluxo de navegação com dados simulados; não implementa um classificador, não foi validado com famílias e não comprova acesso a conteúdo de outros aplicativos. As etapas futuras são: (i) avaliar a viabilidade de um classificador no aparelho; (ii) estudar o aperfeiçoamento do modelo com proteção dos dados locais; (iii) validar a experiência com responsáveis e adolescentes; e (iv) especificar a governança de dados. Os critérios numéricos de aprovação deverão ser definidos antes dos testes confirmatórios, com base em um estudo-piloto, e não por transposição de resultados de outros conjuntos de dados ou dispositivos.

A primeira etapa é determinar se é viável substituir os dados simulados por um classificador executado no aparelho do usuário. A inferência local é uma propriedade da arquitetura, mas não resolve como o aplicativo obteria o texto a analisar. Será necessária uma prova de conceito que identifique a fonte do conteúdo, as permissões exigidas e as restrições técnicas e de privacidade. Uma alternativa de escopo inicial é analisar conteúdo que o próprio usuário compartilhe voluntariamente com o GuardaMirim, sem pressupor leitura irrestrita de outros aplicativos.

Para o português brasileiro, há modelos e conjuntos de dados que podem apoiar a investigação: o BERTimbau, modelo pré-treinado (SOUZA; NOGUEIRA; LOTUFO, 2020); a ToLD-Br, de publicações anotadas para linguagem tóxica, cujo trabalho de referência relata cerca de 76% de macro-F1 em tarefa binária (LEITE et al., 2020); e a HateBR, de comentários do Instagram anotados para ofensividade e discurso de ódio (VARGAS et al., 2022). Esses números servem de referência para comparar resultados publicados, e não como meta transferível ao GuardaMirim.

A avaliação deverá primeiro especificar o que será classificado, pois linguagem ofensiva, discurso de ódio e cyberbullying são fenômenos relacionados, mas não equivalentes. Exigirá categorias operacionais, diretrizes de anotação, dados de teste separados dos de treinamento e relato de macro-F1, precisão e revocação por categoria, com exame de falsos positivos e negativos. A escassez de bases rotuladas representativas é um dos desafios centrais

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

da área (SALAWU; HE; LUMSDEN, 2020), e gírias e ironia de adolescentes exigirão protocolo ético adequado. A execução em dispositivos móveis é outro teste: o MobileBERT reduziu tamanho e latência em benchmarks específicos (SUN et al., 2020), o que não prediz o desempenho de um modelo em português nos aparelhos do público-alvo. Deverão ser medidos tamanho, latência, memória e consumo de energia, com limites fixados após o estudo-piloto.

Outra etapa é avaliar se os botões “Concordo” e “Discordo” da tela de detalhes do alerta (RF10) podem gerar feedback útil para aprimorar o classificador sem enviar a mensagem original a um servidor. Esse feedback não equivale a um rótulo de verdade: “Discordo” indica desacordo, mas não comprova um falso positivo. O aprendizado federado, em que o treinamento ocorre nos dispositivos e apenas atualizações de parâmetros são agregadas (MCMAHAN et al., 2017), é uma alternativa a investigar; já foi aplicado à previsão de texto em teclados móveis (HARD et al., 2018), sem que isso comprove sua adequação ao GuardaMirim. As atualizações de modelo também podem revelar informações indiretas, de modo que “não enviar o texto” não garante privacidade absoluta.

A terceira etapa é avaliar a usabilidade e a compreensão do protótipo com responsáveis e adolescentes (seção 4.6). Estudos sobre aplicativos de controle parental apontam tensão entre supervisão, segurança, autonomia e privacidade dos adolescentes (WISNIEWSKI et al., 2017; GHOSH et al., 2018a), e outro estudo associou o uso de aplicativos de monitoramento a mais experiências de risco online, enquanto o envolvimento direto dos responsáveis se associou a menos problemas; por ser evidência contextual, não deve ser tomada como causal nem generalizada (GHOSH et al., 2018b). Prevê-se um estudo-piloto com os dois perfis, combinando tarefas observadas, entrevistas semiestruturadas e a System Usability Scale (SUS) (BROOKE, 1996). A média de cerca de 70 pontos e o valor de 71,4 associado ao adjetivo “bom” (BANGOR; KORTUM; MILLER, 2009) servem como parâmetros interpretativos, e não como limite universal de aprovação.

A SUS será complementada por indicadores como conclusão correta de tarefas, erros, compreensão do que é e do que não é analisado e percepção de privacidade. Por envolver adolescentes, o plano será submetido à avaliação ética institucional antes do recrutamento, com consentimento dos responsáveis, assentimento dos adolescentes e proteção dos registros; o enquadramento deverá ser confirmado com o Comitê de Ética em Pesquisa, inclusive quanto à Resolução CNS nº 510/2016 (BRASIL, 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

5.3 Governança de dados, privacidade e trabalhos futuros

A quarta etapa é especificar uma política de dados coerente com a arquitetura e com os direitos de crianças e adolescentes. Mesmo sem a transmissão do conteúdo das mensagens, metadados de alerta como categoria, aplicativo de origem e horário podem ser dados pessoais quando relacionados a um menor, e finalidade, necessidade e melhor interesse deverão ser avaliados para cada dado e etapa do tratamento (BRASIL, 2018, art. 14; BRASIL, 2025, arts. 7º e 16).

A Lei nº 15.211/2025 determina que o controlador mapeie e procure mitigar os riscos do tratamento de dados de crianças e adolescentes e elabore relatório de impacto, e o Decreto nº 12.880/2026 trata da avaliação de impacto à segurança e à saúde de crianças (BRASIL, 2025, art. 16; BRASIL, 2026, art. 47). O projeto deverá examinar a incidência dessas exigências sobre seu desenho e concluir a avaliação de riscos antes de qualquer piloto, como medida de governança do próprio trabalho. A arquitetura deverá ainda ser analisada à luz da vedação a mecanismos de vigilância massiva, genérica ou indiscriminada (BRASIL, 2025, art. 34, § 1º, e art. 37, parágrafo único), avaliando quais conteúdos o sistema alcança, quais metadados retém e quem pode consultar alertas.

6 Considerações finais

A presente pesquisa teve como objetivo analisar os mecanismos tecnológicos de proteção de crianças e adolescentes no ambiente digital, considerando os riscos associados ao cyberbullying e a outros comportamentos prejudiciais, bem como as disposições da Lei Geral de Proteção de Dados Pessoais (LGPD) e do Estatuto Digital da Criança e do Adolescente (ECA Digital). A partir da análise realizada, foi possível compreender a importância de conciliar segurança, privacidade e supervisão parental na proteção desse público.

A análise dos mecanismos de segurança das plataformas Discord, Instagram, TikTok e WhatsApp demonstrou que, embora existam recursos de moderação de conteúdo, verificação de idade e controle parental, essas ferramentas apresentam limitações, principalmente na identificação de situações de risco em conversas privadas. Dessa forma, observou-se a necessidade de soluções complementares que auxiliem os responsáveis na prevenção de comportamentos prejudiciais, respeitando a privacidade e a autonomia de crianças e 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

Nesse contexto, foi desenvolvido o protótipo do aplicativo Guarda Mirim, uma proposta tecnológica voltada ao apoio à supervisão parental. O trabalho resultou na definição de requisitos, na elaboração de uma arquitetura e na implementação de sete telas navegáveis, com funcionalidades voltadas à apresentação de alertas e orientações aos responsáveis. A proposta busca permitir a identificação de situações de risco sem expor o conteúdo original das conversas, priorizando os princípios de segurança e proteção de dados.

Entretanto, é importante destacar que o protótipo utiliza dados simulados e ainda não possui mecanismos automatizados de detecção implementados. Além disso, não foram realizados testes com famílias ou avaliações práticas de sua eficácia. Assim, a hipótese apresentada foi sustentada no âmbito conceitual, mas ainda depende de validação. Como continuidade da pesquisa, recomenda-se investigar a viabilidade da classificação automatizada de conteúdos, realizar testes de usabilidade com responsáveis e adolescentes e aprofundar os mecanismos de privacidade e proteção de dados.

Ressalta-se, por fim, que a tecnologia não substitui o diálogo familiar e a educação digital, mas pode atuar como uma ferramenta de apoio à construção de um ambiente virtual mais seguro.

Referências

AGÊNCIA BRASIL. ECA Digital começa a valer nesta terça; confira principais pontos. Brasília, 17 mar. 2026. Disponível em: https://agenciabrasil.ebc.com.br/politica/noticia/2026-03/eca-digital-comeca-valer-nesta-terca-confira-principais-pontos. Acesso em: 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. Disponível em: 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. Acesso em: 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. Disponível em: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Bangor_May2009.pdf. Acesso em: 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. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/decreto/d12880.htm. Acesso em: 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. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13185.htm. Acesso em: 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. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 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. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/lei/l14811.htm. Acesso em: 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. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/l15211.htm. Acesso em: 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. Disponível em: https://canaltech.com.br/apps/whatsapp-ganha-novos-controles-para-pais-de-adolescentes-veja-o-que-muda/. Acesso em: 1 nov. 2026.

DISCORD. Discord Family Center. [S. l.], [2026]. Disponível em: https://discord.com/safety-family-center. Acesso em: 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. Disponível em: https://jornaldebrasilia.com.br/noticias/brasil/morte-de-adolescente-de-13-anos-leva-policia-a-grupo-suspeito-de-induzir-jovens-ao-suicidio-pela-internet/. Acesso em: 1 nov. 2026.

FEDERAL TRADE COMMISSION. Children’s Online Privacy Protection Rule (COPPA). Washington, DC, 2025. Disponível em: https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa. Acesso em: 18 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

FUNDAÇÃO ABRINQ PELOS DIREITOS DA CRIANÇA E DO ADOLESCENTE. ECA Digital: o que muda para crianças e adolescentes. São Paulo, 6 jul. 2026. Disponível em: https://www.fadc.org.br/noticias/eca-digital-o-que-muda. Acesso em: 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. Disponível em: https://arxiv.org/abs/1811.03604. Acesso em: 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. Disponível em: https://arxiv.org/abs/2010.04543. Acesso em: 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. Disponível em: https://about.fb.com/br/news/2026/03/apresentamos-as-contas-gerenciadas-por-pais-no-whatsapp/. Acesso em: 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). Disponível em: https://about.fb.com/br/news/2026/06/iniciativas-para-manter-pais-brasileiros-

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

informados-sobre-as-contas-de-adolescente-e-supervisao-parental-no-instagram/. Acesso em: 1 nov. 2026.

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

MIGALHAS. Governo processa Discord, pede R$ 500 mi e cobra proteção a menores. [S. l.], 26 ago. 2026. Disponível em: https://www.migalhas.com.br/quentes/463309/governo-processa-discord-pede-r-500-mi-e-cobra-protecao-a-menores. Acesso em: 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. Disponível em: https://piauihoje.com/index.php/noticias/brasil/mpf-investiga-seguranca-do-discord-para-proteger-menores-463658.html. Acesso em: 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. Disponível em: https://www.showmetech.com.br/governo-federal-proibe-o-discord-para-menores-de-18-anos/. Acesso em: 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. Disponível em: https://newsroom.tiktok.com/nosso-compromisso-inequvoco-com-a-segurana-dos-jovens-no-tiktok?lang=pt-BR. Acesso em: 1 nov. 2026.

TIKTOK. Privacidade e segurança para usuários menores de 18 anos. [S. l.], [2026?]. Disponível em: https://support.tiktok.com/pt_BR/account-and-privacy/account-privacy-settings/privacy-and-safety-settings-for-users-under-age-18. Acesso em: 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 tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva

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

95/46/CE (Regulamento Geral sobre a Proteção de Dados). Jornal Oficial da União Europeia, Luxemburgo, L 119, p. 1-88, 4 maio 2016. Disponível em: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Acesso em: 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. Disponível em: https://aclanthology.org/2022.lrec-1.777. Acesso em: 18 out. 2026.

WHATSAPP. Como manter a segurança no WhatsApp. [S. l.], [2026?]. Disponível em: https://faq.whatsapp.com/1313491802751163/?locale=pt_BR. Acesso em: 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.