



## ALERTAS PARENTAIS E DETECÇÃO ALGORÍTMICA DE RISCO DE AUTOAGRESSÃO EM REDES SOCIAIS: PROPORCIONALIDADE, LGPD E ECA DIGITAL A PARTIR DO CASO META/INSTAGRAM (2026)

### *PARENTAL ALERTS AND ALGORITHMIC DETECTION OF SELF-HARM RISK ON SOCIAL NETWORKS: PROPORTIONALITY, BRAZIL'S LGPD AND THE DIGITAL CHILD AND ADOLESCENT STATUTE FROM THE META/INSTAGRAM (2026) CASE*

**LUCIANO ROCHA DE OLIVEIRA**

Faculdade de Direito de Vitória. Mestre e Doutorando em Direitos e Garantias Fundamentais na FDV/ES, membro do Grupo de Pesquisa Hermenêutica e Jurisdição Constitucional. Promotor de Justiça. Vitória/ES, Brasil. E-mail: [lucianompes@gmail.com](mailto:lucianompes@gmail.com)

**AMÉRICO BEDÊ FREIRE JÚNIOR**

Faculdade de Direito de Vitória. Doutor e Mestre em Direitos Fundamentais. Professor do Programa de Mestrado e Doutorado da FDV/ES, coordenador do Grupo de Pesquisa Hermenêutica e Jurisdição Constitucional. Juiz Federal. Vitória/ES, Brasil. E-mail: [bedejunior@hotmail.com](mailto:bedejunior@hotmail.com)

#### RESUMO:

O artigo analisa a compatibilidade de alertas parentais, baseados na detecção automatizada de risco de autoagressão em redes sociais, com a LGPD e com o ECA Digital (Lei 15.211/25), à luz do princípio da proporcionalidade, tomando como referência o caso Meta/Instagram (2026). Sustenta-se que, embora a medida possa ter finalidade protetiva legítima, sua validade jurídico-constitucional depende do modo como é desenhada e operada: transparência significativa para o adolescente e os responsáveis, minimização e segurança dos dados, limites claros de acionamento, revisão periódica do mecanismo, suporte emocional prioritário ao adolescente e instrumentos efetivos de contestação. Defende-se, por fim, que alertas parentais, sem salvaguardas robustas, podem produzir efeitos contraproducentes, como violação desproporcional da privacidade e erosão da autonomia progressiva, com potencial de estigmatização e de retração na busca por ajuda.

**Pesquisa:** Em que condições alertas parentais baseados na detecção automatizada do risco de autoagressão em redes sociais podem ser considerados proporcionais e compatíveis com a LGPD e com o ECA Digital, sem produzir efeitos contraproducentes sobre a privacidade, a autonomia progressiva e a segurança de adolescentes?





**Metodologia:** Realiza-se pesquisa jurídico-dogmática, com análise documental (comunicado oficial da Meta e normas brasileiras pertinentes), revisão bibliográfica sobre governança algorítmica, explicabilidade e proteção de crianças e adolescentes no ambiente digital, além da aplicação estruturada do teste de proporcionalidade.

**Resultados:** Identifica-se que a medida pode ser juridicamente defensável como mecanismo protetivo somente quando acompanhada de salvaguardas robustas de transparência, minimização, segurança da comunicação, revisão periódica do mecanismo e suporte emocional, sob pena de violação desproporcional à privacidade e à autonomia progressiva do adolescente.

**Contribuições:** Propõe-se um conjunto de diretrizes de conformidade (*by design*) para alertas parentais em plataformas digitais, articulando o ECA Digital e a LGPD com a governança e a contestabilidade de decisões automatizadas.

**Aplicações Práticas e Normativas:** As diretrizes sistematizadas podem orientar programas de conformidade e desenho de produto em plataformas digitais, servindo também como parâmetros de avaliação regulatória e de controle (em especial quanto à transparência, minimização, segurança, revisão/contestação e prevenção de danos). Em termos normativos, contribuem para delimitar quando a intervenção protetiva é excepcional e justificável, evitando que a proteção integral seja operacionalizada como monitoramento amplo e potencialmente desproporcional.

**Palavras-chave:** Crianças e adolescentes; ECA Digital; LGPD; redes sociais; detecção algorítmica de autoagressão.

## ABSTRACT:

*This article examines whether—and under which conditions—parental alerts triggered by automated detection of self-harm risk on social networks can be compatible with Brazil’s LGPD and the Digital Child and Adolescent Statute, using a proportionality analysis and the Meta/Instagram (2026) case as a reference point. The study argues that, although the measure may pursue a legitimate protective aim, its legal and constitutional acceptability depends on strict safeguards: meaningful transparency for adolescents and guardians, data minimization and security, clear triggering limits, periodic review of the mechanism, priority access to support resources for the adolescent, and effective avenues for contestation. Without such safeguards, parental alerts may generate counterproductive outcomes, including disproportionate interference with privacy and the adolescent’s evolving capacities, as well as risks of stigmatization and reduced help-seeking behavior.*

**Research:** *Under which conditions can parental alerts based on automated detection of self-harm risk on social networks be considered proportionate and compatible with Brazil’s LGPD and the Digital Child and Adolescent Statute, without producing counterproductive effects on adolescents’ privacy, evolving autonomy, and safety?*

**Methodology:** *The article adopts a legal-dogmatic approach, combining documentary analysis (Meta’s official communications and relevant Brazilian norms), a bibliographic review of algorithmic governance, explainability, and the protection of children and adolescents in digital environments, and a structured application of the proportionality test.*

**Results:** *The study finds that the measure may be legally defensible as a protective mechanism only if accompanied by robust safeguards—transparency, minimization,*





secure communication channels, periodic review, and support resources—otherwise it risks disproportionate violations of privacy and evolving autonomy.

**Contributions:** The article proposes a set of “by design” compliance guidelines for parental alerts on digital platforms, bridging the LGPD and the Digital Child and Adolescent Statute, and addressing governance and contestability requirements for automated decision-making.

**Practical and Normative Applications:** The proposed guidelines can support platform compliance and product design, and provide evaluative parameters for regulatory oversight, especially regarding transparency, minimization, security, periodic review, and effective contestation. Normatively, they help define when a protective intervention is exceptional and justifiable, preventing “integral protection” from being implemented in the form of broad, potentially disproportionate monitoring.

**Keywords:** Children and adolescents; Digital Child and Adolescent Statute; LGPD; social networks; algorithmic detection of self-harm risk.

## 1 INTRODUÇÃO

Em fevereiro de 2026, a Meta anunciou a implementação, no Instagram, de um mecanismo de alertas parentais orientado a situações de risco potencial: quando um adolescente tentar, em curto período, pesquisar repetidamente termos associados a suicídio ou automutilação, pais ou responsáveis que utilizem as ferramentas de supervisão receberão notificações acompanhadas de acesso a recursos de apoio especializado. A própria empresa também sinaliza a intenção de desenvolver notificações semelhantes relacionadas às interações de adolescentes com ferramentas de IA, aproximando o tema da agenda contemporânea de detecção automatizada de risco em ambientes digitais (META, 2026b).

A proposta, embora orientada por finalidade protetiva, explicita um dilema jurídico estrutural. De um lado, há um dever reforçado – compartilhado por família, sociedade e Estado e também projetado sobre atores privados que operam serviços digitais – de reduzir riscos previsíveis à saúde e à vida de crianças e adolescentes. De outro lado, a arquitetura do alerta pressupõe monitoramento de padrões de busca e processamento automatizado de dados, com potencial de inferência sobre vulnerabilidade, intensificando assimetrias informacionais e tensionando privacidade, autodeterminação informativa e autonomia progressiva. Em termos de cidadania digital, o problema não é optar entre proteção e privacidade, mas construir critérios jurídicos para compatibilizá-las de modo racional, verificável e controlável, sobretudo





quando a triagem é automatizada (CANTARINI, 2023, p. 774; CORRADINI et al., 2023, p. 348).

No contexto brasileiro, essa compatibilização exige leitura integrada de marcos normativos que incidem diretamente sobre o desenho de produtos digitais. A Lei Geral de Proteção de Dados Pessoais estabelece parâmetros materiais e procedimentais para o tratamento de dados de crianças e adolescentes, com centralidade no melhor interesse e garantias associadas à transparência, à finalidade, à necessidade e à revisão de decisões tomadas exclusivamente com base em tratamento automatizado (BRASIL, 2018). O Estatuto Digital da Criança e do Adolescente, por sua vez, reforça obrigações de segurança, prevenção e arquitetura protetiva proporcional, com ênfase em medidas adequadas para assegurar alto nível de privacidade e segurança no desenho e na operação de produtos e serviços acessados por esse público (BRASIL, 2025). Partindo desse enquadramento e tomando o caso Meta/Instagram (2026) como referência empírica de governança algorítmica protetiva, a análise aplica o teste de proporcionalidade, em sua estrutura clássica de adequação, necessidade e proporcionalidade em sentido estrito, conforme sistematizado por Alexy (2015, p. 116–120), para responder à seguinte pergunta: Em que condições alertas parentais baseados na detecção automatizada do risco de autoagressão em redes sociais podem ser considerados proporcionais e compatíveis com a LGPD e com o ECA Digital, sem produzir efeitos contraproducentes sobre a privacidade, a autonomia progressiva e a segurança de adolescentes?

## 2 O CASO META/INSTAGRAM (2026): DESCRIÇÃO DO MECANISMO DE ALERTA PARENTAL E SEU CONTEXTO

Esta seção delimita, de forma estritamente documental, o objeto empírico do estudo: o que a Meta afirmou publicamente sobre a implementação de alertas parentais vinculados a tentativas de busca associadas a suicídio ou automutilação no Instagram, distinguindo a descrição fática dos juízos normativos.

Em 26 de fevereiro de 2026, a Meta anunciou que passará a implementar alertas parentais acionados por atividades de pesquisa realizadas por adolescentes na plataforma, no contexto das Contas de Adolescente (*teen accounts*) e das ferramentas de supervisão parental (META, 2026b). Para fins deste artigo, denomina-





se “caso Meta/Instagram (2026)” o arranjo composto pelas regras de acionamento do alerta, pelas condições de habilitação, pelos canais de comunicação aos responsáveis e pela expansão anunciada para experiências envolvendo IA, conforme descrito pela própria empresa (META, 2026b). O estudo não assume, como premissa, a existência de uma ordem judicial específica que imponha a mudança, mas trabalha apenas com o que está afirmado nas fontes primárias.

O comunicado informa que os alertas serão enviados aos pais e responsáveis que utilizam as ferramentas de supervisão parental do Instagram, indicando uma lógica de ativação prévia do recurso e não de monitoramento indistinto (META, 2026b). Em texto institucional anterior, a empresa descreve as Contas de Adolescente como um modelo de conta com proteções integradas e possibilidade de definição de limites via supervisão parental, contexto que auxilia a situar o alerta como componente dessa arquitetura de parentalidade digital (META, 2026a).

No plano operacional, a Meta afirma que o alerta será acionado quando o adolescente tentar pesquisar repetidamente, em curto período, termos relacionados a suicídio ou automutilação, incluindo termos explícitos e frases que promovam ou indiquem intenção de autoagressão (META, 2026b). As fontes públicas não detalham se a triagem decorre de listas de termos, de classificação semântica, de modelos de aprendizado de máquina ou de combinação de métodos; por rigor metodológico, o mecanismo é tratado aqui apenas como detecção automatizada baseada em atividade de pesquisa e em critérios publicamente descritos (META, 2026b). A empresa reconhece o risco de “supernotificação” e declara ter calibrado um limiar para evitar comunicações desnecessárias, admitindo que, por cautela, poderá notificar pais mesmo quando “talvez não haja motivo real” para preocupação (META, 2026b).

Quanto à forma de comunicação, o alerta poderá chegar por e-mail, mensagem de texto ou *WhatsApp*, além de notificação no aplicativo; ao acessar a notificação, os responsáveis visualizarão uma mensagem que informa o padrão de busca e terão a opção de acessar recursos especializados para apoiar conversas sensíveis com o adolescente (META, 2026b). O anúncio descreve a implantação em fases, com lançamento inicial para supervisão parental nos Estados Unidos, Reino Unido, Austrália e Canadá, e previsão de expansão para outras regiões ainda em 2026 (META, 2026b).

No mesmo comunicado, a Meta sustenta que os alertas se baseiam em políticas e proteções já existentes, como restrições a conteúdos que promovam ou





glorifiquem a autoagressão, ocultação de determinados conteúdos de adolescentes, bloqueio de pesquisas por termos claramente associados e redirecionamento de recursos de apoio (META, 2026b). A empresa também anuncia a intenção de criar alertas semelhantes para certas experiências de IA, notificando pais se um adolescente tentar se envolver em conversas relacionadas a suicídio ou automutilação com a IA da empresa (META, 2026b). Embora a avaliação jurídica seja feita sob a LGPD e o ECA Digital, o anúncio foi noticiado em meio a pressões e a debates públicos sobre a segurança de adolescentes em plataformas, o que reforça a atualidade do tema e a necessidade de um escrutínio normativo cuidadoso (KERR, 2026; REUTERS, 2026).

Em síntese, o caso consiste na observação de um padrão de busca associado a autoagressão, no acionamento de um limiar em curto período (com admissão de falsos positivos), no envio de alerta a responsáveis em regime de supervisão e na oferta de recursos de apoio, com expansão anunciada para interações envolvendo IA; esses elementos delimitam o objeto a ser submetido, adiante, ao enquadramento jurídico e ao teste de proporcionalidade (META, 2026b).

### **3 DIREITOS DA CIDADANIA DIGITAL DE CRIANÇAS E ADOLESCENTES: PROTEÇÃO INTEGRAL, MELHOR INTERESSE E AUTONOMIA PROGRESSIVA**

A avaliação de alertas parentais acionados por detecção automatizada de risco nas redes sociais exige um enquadramento normativo prévio ao debate sobre a técnica algorítmica e a arquitetura de produto. No Direito da Criança e do Adolescente (BRASIL, 1990), a condição peculiar da pessoa em desenvolvimento e a proteção integral orientam os deveres estatais e privados também no ambiente digital. Isso implica reconhecer a legitimidade de medidas voltadas a reduzir riscos extremos, como a autoagressão e o suicídio, sem esvaziar a cidadania digital do adolescente, sobretudo nas dimensões de privacidade, participação e autonomia progressiva. Nessa moldura, o caso Meta/Instagram (2026) deve ser compreendido como uma intervenção protetiva que apenas se sustenta se compatibilizada com as garantias e salvaguardas próprias do regime jurídico de proteção infantojuvenil.

#### **3.1 PROTEÇÃO INTEGRAL E PRIORIDADE ABSOLUTA COMO DEVERES**





## DIRIGIDOS TAMBÉM AO ECOSISTEMA DIGITAL

O ponto de partida constitucional é o art. 227 da Constituição, que estabelece o dever compartilhado da família, da sociedade e do Estado de assegurar, com absoluta prioridade, os direitos fundamentais de crianças, adolescentes e jovens e de colocá-los a salvo de negligência, discriminação, exploração e violência (BRASIL, 1988). O Estatuto da Criança e do Adolescente densifica essa diretriz ao estruturar a doutrina da proteção integral e ao reafirmar a prioridade na efetivação dos direitos (BRASIL, 1990). No ambiente digital, esse eixo não se limita a políticas públicas tradicionais: projeta expectativas normativas sobre arranjos privados que hoje organizam parte relevante da vida social de adolescentes, como as plataformas de redes sociais. É nessa direção que se insere o Estatuto Digital da Criança e do Adolescente, ao adotar uma lógica funcional e de risco para alcançar produtos e serviços digitais direcionados a crianças e adolescentes ou de acesso provável por eles, especialmente quando envolvam risco à privacidade, à segurança e ao desenvolvimento biopsicossocial (BRASIL, 2025). Com isso, os alertas parentais deixam de ser vistos como simples funcionalidade e passam a ser compreendidos como prática ligada ao dever de prevenção e mitigação de riscos psicossociais em ambiente digital, sujeita a parâmetros de legalidade, necessidade e proporcionalidade (BRASIL, 2025).

### 3.2 MELHOR INTERESSE COMO CRITÉRIO SUBSTANTIVO, PRINCÍPIO INTERPRETATIVO E REGRA PROCEDIMENTAL NO AMBIENTE DIGITAL

O princípio do melhor interesse atua como critério decisório quando há tensão entre valores e direitos no sistema de proteção de crianças e adolescentes. Segundo o Comitê sobre os Direitos da Criança, o *best interests of the child* deve ser entendido como direito substantivo, princípio interpretativo e regra de procedimento, exigindo que as decisões que afetem crianças e adolescentes sejam justificadas por razões explícitas e controláveis (UNITED NATIONS, 2013). No ambiente digital, essa dimensão procedimental é particularmente sensível: intervenções protetivas baseadas em sinais comportamentais podem ser bem-intencionadas e, ainda assim, produzir efeitos contraproducentes se operarem com baixa transparência, critérios opacos ou sem revisão e correção.





O Estatuto Digital da Criança e do Adolescente internaliza esse vetor ao exigir que produtos e serviços voltados a crianças e adolescentes, ou de acesso provável por eles, garantam proteção prioritária e adotem medidas adequadas e proporcionais para assegurar elevado nível de privacidade, proteção de dados e segurança, tomando o melhor interesse como parâmetro (BRASIL, 2025). Além disso, ao reconhecer que a expressão do melhor interesse compreende, simultaneamente, privacidade, segurança, saúde mental e física, acesso à informação, participação, acesso significativo às tecnologias digitais e bem-estar, a lei impede leituras simplificadoras segundo as quais proteger equivaleria sempre a ampliar a supervisão e a notificação (BRASIL, 2025). Em mecanismos como o alerta parental, a consequência prática é direta: se a intervenção pretende prevenir dano extremo, precisa demonstrar, em critérios e salvaguardas, por que a comunicação é necessária e como se evitam abusos, erros e escaladas familiares, sob pena de o melhor interesse ser invocado apenas de forma retórica.

Autonomia progressiva, privacidade e participação: limites internos da supervisão parental em nome da cidadania digital

A proteção integral não autoriza a substituição completa da esfera privada do adolescente pelo monitoramento. A noção de autonomia progressiva impõe graduar as intervenções conforme a idade, a maturidade e o contexto, reconhecendo que o exercício de direitos no ambiente digital integra o próprio desenvolvimento. Nessa linha, o ECA Digital incorpora, entre seus fundamentos, o respeito à autonomia e ao desenvolvimento progressivo, bem como a exigência de transparência e responsabilidade no tratamento de dados pessoais (BRASIL, 2025). Em convergência, o Comitê sobre os Direitos da Criança recomenda que os Estados promovam alfabetização digital e orientações a pais e cuidadores, mas ressalta a necessidade de respeitar a autonomia evolutiva e a privacidade de crianças e adolescentes, para evitar que a proteção se converta em interferência indevida (UNITED NATIONS, 2021). No plano nacional, a Resolução CONANDA nº 257/2024 reforça diretrizes de proteção no ambiente digital, incluindo autonomia progressiva, escuta e participação de crianças e adolescentes, além da proteção de dados e da privacidade desde a concepção e por padrão, com responsabilidade compartilhada que abrange empresas provedoras de produtos e serviços digitais (BRASIL, 2024). Esse conjunto normativo confere densidade à ideia de que a supervisão parental, embora legítima, deve operar com transparência, proporcionalidade interna e





salvaguardas capazes de evitar a transformação do cuidado em controle permanente.

Quando a intervenção se apoia em sistemas de IA, a exigência de limites e garantias se intensifica. Diretrizes internacionais voltadas à IA centrada em crianças enfatizam privacidade, segurança, transparência e *accountability* em sistemas capazes de inferir estados emocionais ou classificar riscos, justamente porque erros, vieses e opacidade tendem a afetar de forma mais severa públicos vulneráveis em desenvolvimento (UNICEF, 2021, 2025). Assim, a cidadania digital de crianças e adolescentes deve ser compreendida como articulação entre proteção integral e melhor interesse, de um lado, e autonomia progressiva, privacidade e participação, de outro, fornecendo o critério normativo para avaliar, nas seções seguintes, se alertas parentais por detecção automatizada de risco podem ser concebidos como medida protetiva proporcional ou se representam interferência excessiva em direitos fundamentais do adolescente.

## 4 ENQUADRAMENTO JURÍDICO DO ALERTA PARENTAL COMO TRATAMENTO DE DADOS

O alerta parental descrito no caso Meta/Instagram (2026) deve ser compreendido, antes de tudo, como uma intervenção protetiva baseada em tratamento de dados: a plataforma observa a atividade de busca do adolescente, aplica um critério de acionamento (“buscas repetidas em curto período”) e, se atingido um limiar, comunica aos responsáveis uma informação sobre possível necessidade de suporte, acompanhada de recursos de orientação (META, 2026b). Ainda que o discurso empresarial se apresente como medida de segurança e cuidado, o enquadramento jurídico não pode se limitar à presunção de boa-fé, pois o próprio desenho do mecanismo envolve a coleta, análise, classificação e comunicação de informações pessoais potencialmente íntimas. Por isso, a legitimidade do instrumento, no ordenamento brasileiro, depende de uma leitura combinada da LGPD (BRASIL, 2018) e do Estatuto Digital da Criança e do Adolescente (BRASIL, 2025), especialmente quanto à minimização, à transparência, à contestabilidade e à proteção desde a concepção.

### 4.1 LGPD: ART. 14, DADOS SENSÍVEIS POR INFERÊNCIA E PRINCÍPIOS DA





## FINALIDADE, DA NECESSIDADE, DA TRANSPARÊNCIA E DA SEGURANÇA

No plano conceitual, tentativas de busca realizadas por adolescente autenticado em uma conta relacionam-se ao titular identificável e, portanto, configuram dado pessoal; do mesmo modo, o disparo (ou não) do alerta e o envio de mensagem aos responsáveis integram o tratamento, por envolverem processamento e comunicação de informação (BRASIL, 2018). O caso adquire especial sensibilidade quando se considera que o próprio comunicado da Meta delimita o gatilho do alerta a pesquisas associadas ao suicídio e à automutilação, inclusive por termos explícitos e por frases que indiquem intenção de se ferir (META, 2026b). Não é necessário afirmar que toda busca sobre o tema revela a saúde do adolescente; basta reconhecer que, em cenários de triagem e classificação de risco, há uma zona de contato entre dados comportamentais e inferências potencialmente sensíveis, o que recomenda cautela reforçada, inclusive porque o ECA Digital explicita que a expressão do melhor interesse abrange a proteção da saúde mental de crianças e adolescentes (BRASIL, 2025).

Do ponto de vista normativo, essa cautela se traduz, primeiro, na incidência dos princípios da LGPD: o tratamento deve observar boa-fé e finalidade, adequação, necessidade, transparência, segurança e prevenção, exigindo coerência entre a finalidade protetiva declarada e o desenho do tratamento (BRASIL, 2018). Em um mecanismo de alerta parental, isso impõe minimizar tanto o gatilho quanto o conteúdo comunicado; transparência sobre o funcionamento geral e as consequências possíveis; e medidas técnicas e organizacionais que reduzam o risco de exposição indevida de mensagens capazes de gerar estigma ou conflitos familiares (BRASIL, 2018). Em segundo lugar, incide a camada específica de crianças e adolescentes: o art. 14 determina que o tratamento seja realizado em seu melhor interesse e disciplina o consentimento parental para crianças, com exceções restritas e limites quanto ao armazenamento e ao compartilhamento (BRASIL, 2018). Ainda que o caso Meta/Instagram (2026) seja apresentado como medida voltada a *teen accounts* e à supervisão parental, o art. 14 reforça que a proteção não é apenas retórica, mas um critério substantivo e transversal de controle de legitimidade, sobretudo se o desenho alcançar crianças por falhas de verificação etária ou por expansão do recurso.

Por fim, se a classificação de risco se aproximar de um dado sensível por inferência, o art. 11 ganha centralidade, inclusive por prever hipóteses específicas que





podem ser invocadas em cenários de proteção da vida ou da incolumidade física do titular ou de terceiro (BRASIL, 2018). O ponto não é fixar, de antemão, uma base legal ideal, mas registrar que medidas protetivas baseadas em sinais de autoagressão devem ser justificadas em termos de necessidade e proporcionalidade e, quando envolvem inferências sensíveis, exigem uma justificativa mais robusta, para evitar que a finalidade protetiva funcione como atalho para ampliar o monitoramento ou o compartilhamento com baixa governança.

## 4.2 LGPD: DECISÕES AUTOMATIZADAS E CONTESTABILIDADE (ART. 20)

A qualificação do alerta como decisão automatizada não é automática. A Meta descreve o acionamento como “buscas repetidas em curto período”, mas não detalha se há intervenção humana em cada caso, se há supervisão posterior ou se o disparo é integralmente automatizado (META, 2026b). Ainda assim, o art. 20 da LGPD (BRASIL, 2018) funciona como parâmetro interpretativo relevante: há um critério operacional de limiar e um *output* (alerta) com potencial de produzir consequências relevantes na esfera privada do adolescente, e a própria empresa reconhece o risco de falsos positivos, o que reforça a necessidade de contestabilidade e de correção. Mesmo que se sustente que o art. 20 não incide diretamente em todos os cenários; seu conteúdo normativo (revisão humana, informação sobre critérios, auditoria) fornece parâmetros de desenho institucional para intervenções protetivas baseadas em triagem automatizada. Nesse ponto, é importante manter coerência com a literatura brasileira que aborda criticamente as decisões automatizadas e o alcance do art. 20, evitando prometer mais do que a lei assegura; Corradini et al. (2023, p. 351) recomendam coerência argumentativa ao utilizar o dispositivo como eixo de contestabilidade no presente tema.

Além disso, o ECA Digital reforça uma orientação de devido processo nas decisões de plataforma, ao menos no campo da retirada de conteúdo, exigindo notificação, motivo e fundamentação, informação sobre se a identificação decorreu de análise humana ou automatizada, além de recurso e prazos (BRASIL, 2025). Ainda que esse comando se refira a procedimento de retirada de conteúdo, ele sinaliza uma opção legislativa por contestabilidade e transparência quanto ao uso de automação, o que, por coerência sistemática, fortalece a exigência de mecanismos de informação e revisão também em alertas parentais baseados em tratamento automatizado.





## 4.3 ECA DIGITAL: SUPERVISÃO PARENTAL, REVISÃO DE IA E SUPORTE EMOCIONAL

O Estatuto Digital da Criança e do Adolescente fornece a ponte normativa entre o dever de proteção e o desenho técnico de funcionalidades em plataformas digitais. No art. 6º impõe que fornecedores adotem medidas razoáveis desde a concepção e ao longo da operação para prevenir e mitigar riscos, incluindo, expressamente, práticas que levem a danos à saúde física ou mental, com menção à automutilação e ao suicídio (BRASIL, 2025). A finalidade protetiva, portanto, é legítima no ecossistema regulatório; porém, não opera como “carta branca”, pois a mesma arquitetura normativa atrela a prevenção à proteção por padrão, transparência e respeito à autonomia e ao desenvolvimento progressivo. Essa lógica se manifesta com clareza no art. 7º: produtos e serviços devem operar, por padrão, com o grau mais elevado de proteção e com informações claras para escolhas informadas sobre configurações menos protetivas (BRASIL, 2025), o que, no caso de alertas parentais, reforça a necessidade de minimizar o conteúdo comunicado e oferecer aviso claro e compreensível ao adolescente.

No capítulo de supervisão parental, o art. 17 exige configurações acessíveis e fáceis de usar, além de um aviso claro e visível quando as ferramentas estiverem em vigor, indicando quais controles foram aplicados (BRASIL, 2025). O comunicado da Meta afirma que pais e adolescentes inscritos em supervisão serão notificados sobre o envio de alertas com base na atividade de busca, o que é compatível, ao menos em tese, com esse dever de informação (BRASIL, 2025; META, 2026b). O ponto mais sensível para a agenda de IA e saúde mental é o §4º do art. 17, ao exigir o mais alto nível de proteção por padrão e, de modo especialmente relevante, revisão regular das ferramentas de inteligência artificial com participação de especialistas e órgãos competentes, bem como a disponibilização, quando tecnicamente viável, de recursos ou conexões a suporte emocional e bem-estar, com orientações baseadas em evidências, especialmente em interações com riscos psicossociais identificados (BRASIL, 2025). A Meta declara ter consultado especialistas para calibrar o limiar de alertas e afirma que os alertas virão acompanhados de recursos especializados, além de anunciar a expansão de experiências de IA relacionadas à autoagressão (META, 2026b), de modo que o parâmetro de conformidade passa a abranger não apenas a existência do alerta, mas também sua governança e revisão contínua.





Por fim, o art. 16 impõe deveres de informação sobre riscos e medidas de segurança, incluindo privacidade e proteção de dados, em conformidade com o art. 14 da LGPD (BRASIL, 2018), e prevê obrigações, como o mapeamento de riscos, esforços de mitigação, elaboração de relatório de impacto, monitoramento e avaliação (BRASIL, 2025). Somado a isso, o Estatuto contém diretrizes de contenção contra derivas de vigilância: a regulamentação não poderá autorizar ou resultar na implantação de mecanismos de vigilância massiva, genérica ou indiscriminada, nem de práticas contrárias a direitos fundamentais, inclusive à privacidade e ao tratamento diferenciado de dados de crianças e adolescentes; também veda a criação de perfis comportamentais para direcionamento de publicidade comercial (BRASIL, 2025). Esses comandos fecham o círculo normativo do caso: medidas protetivas podem existir e até ser exigidas, mas devem ser desenhadas para não se converterem em infraestrutura de monitoramento amplo, reutilização de dados e perfilamento.

## 5 TESTE DE PROPORCIONALIDADE: ADEQUAÇÃO, NECESSIDADE E PROPORCIONALIDADE EM SENTIDO ESTRITO

A tensão central do caso Meta/Instagram (2026) pode ser descrita, em termos jurídico-constitucionais, como um conflito entre uma finalidade protetiva de altíssima relevância, qual seja: prevenir danos graves associados ao suicídio e à automutilação, especialmente em adolescentes; e a intensidade de interferência na esfera privada do titular, pois a medida converte um comportamento informacional (padrões de busca) em comunicação a terceiros (pais/responsáveis), com risco reconhecido de supernotificação e impactos familiares imprevisíveis (META, 2026b). Em situações assim, a aplicação do teste de proporcionalidade funciona como método de controle racional da justificabilidade de intervenções que afetam direitos fundamentais, permitindo verificar se a medida é apta, necessária e equilibrada quando comparada à intensidade da restrição que impõe (ALEXY, 2015, p. 116–120). No ordenamento brasileiro, esse exame precisa dialogar de modo sistemático com a LGPD, que impõe princípios como finalidade, adequação, necessidade, transparência, segurança e prestação de contas, e com o ECA Digital, que organiza deveres específicos de prevenção e mitigação de riscos psicossociais no ambiente digital, combinando proteção elevada por padrão, supervisão parental e exigências de governança e





revisão de ferramentas aplicadas a crianças e adolescentes (BRASIL, 2018, 2025).

## 5.1 ADEQUAÇÃO: O ALERTA É CAPAZ DE REDUZIR O RISCO?

No juízo de adequação, pergunta-se se o alerta parental, tal como desenhado, é capaz de contribuir para a finalidade protetiva que declara perseguir. A cadeia causal proposta pela Meta é relativamente clara: se o adolescente insiste em buscas associadas à autoagressão em curto prazo, isso pode sinalizar sofrimento e necessidade de apoio; ao informar responsáveis que aderiram à supervisão parental, aumenta-se a chance de intervenção tempestiva e de encaminhamento a recursos adequados (META, 2026b). Sob esse ângulo, a medida é, em tese, apta a incrementar a proteção, sobretudo porque a finalidade de reduzir o risco de autoagressão conecta-se diretamente ao dever de prevenção e mitigação de riscos psicossociais imposto aos fornecedores de serviços acessíveis a crianças e adolescentes (BRASIL, 2025).

Essa aptidão, contudo, é limitada por fragilidades inerentes ao próprio mecanismo, com consequências relevantes para a leitura jurídica do ganho protetivo prometido. Primeiro, há subdeterminação técnica: a partir das fontes públicas, não se conhecem os critérios exatos nem a forma de validação do limiar (“algumas buscas em curto período”), o que impede concluir com segurança sobre a acurácia, a robustez e os vieses do modelo. Segundo, a própria empresa admite que o desenho, por cautela, pode gerar comunicações indevidas, o que evidencia um risco estrutural de falsos positivos (META, 2026b). Por isso, a adequação não deve ser tratada como um atributo absoluto (“é adequado” ou “não é”), mas como uma condição que exige governança contínua, com monitoramento de desempenho, calibragem de limiares e revisão periódica, sob pena do instrumento perder utilidade protetiva e aumentar interferências injustificadas. Em termos de LGPD, a própria adequação passa pelo vínculo entre a finalidade declarada e o desenho do tratamento: quanto mais a operação se afastar do estritamente necessário à proteção, mais difícil será sustentar a compatibilidade com os princípios de finalidade e adequação (BRASIL, 2018).

## 5.2 NECESSIDADE: HAVERIA UMA ALTERNATIVA MENOS INTRUSIVA COM EFICÁCIA EQUIVALENTE?

No juízo de necessidade, o teste exige verificar se há meio menos restritivo à





privacidade e à autonomia do adolescente, capaz de assegurar proteção equivalente. Aqui, o alerta parental tem caráter ambivalente. Por um lado, ele é mais intrusivo do que intervenções internas não comunicacionais, como fricções na busca, direcionamento de recursos, mensagens de acolhimento e canais de ajuda apresentados diretamente ao adolescente. Por outro lado, ele pode ser menos intrusivo do que modelos de supervisão ampla e permanente, pois, segundo a descrição pública, opera como um acionamento excepcional quando um padrão de buscas se repete em curto período e se limita a famílias que optaram pela supervisão (META, 2026b). A necessidade, portanto, depende do desenho concreto: quanto mais o sistema for escalonado, priorizando o suporte direto ao adolescente e reservando a comunicação aos responsáveis para casos persistentes e sinais de maior gravidade, mais se aproxima de uma intervenção menos restritiva e potencialmente justificável.

Esse ponto se torna ainda mais exigente quando se consideram alternativas estruturais de proteção por *design*. Auditorias independentes sobre recursos de segurança para adolescentes apontam que parte central do risco não está apenas no que jovens procuram, mas também no que sistemas e interfaces recomendam e tornam “descobrível” para eles, inclusive em contextos em que controles deveriam restringir o acesso ao sensível (BÉJAR, 2025). Sob a lógica do ECA Digital, isso reforça que mecanismos protetivos não podem deslocar o ônus da segurança para a família como solução substitutiva de correções estruturais: a obrigação de prevenir e mitigar riscos desde a concepção e ao longo da operação é do fornecedor, e envolve controles sobre recomendação e restrições a mecanismos de uso compulsivo, além da revisão regular de ferramentas de inteligência artificial e da oferta, sempre que tecnicamente viável, de conexões a suporte emocional, com orientações baseadas em evidências, especialmente em interações com riscos psicossociais identificados (BRASIL, 2025).

Sob a LGPD, o princípio da necessidade opera como exigência de minimização: tratar e comunicar apenas o indispensável ao fim protetivo (BRASIL, 2018). Isso tem implicações diretas para os alertas parentais. Mesmo que se aceite que comunicar a responsáveis pode ser necessário em certos cenários, não se segue que seja necessário comunicar detalhes de termos pesquisados, horários ou histórico amplo; em regra, basta informar que houve repetição de tentativas de busca associadas a risco, vinculando a comunicação a recursos de orientação e canais de ajuda, de modo a reduzir estigmatização e efeitos adversos sem sacrificar o objetivo





de cuidado. Também se torna relevante limitar a retenção e circulação interna do dado que aciona o alerta, sobretudo quando a informação pode carregar inferências sensíveis sobre saúde mental, pois, nesses cenários, a carga justificativa aumenta e o risco de dano em caso de vazamento ou uso indevido é mais elevado (BRASIL, 2018).

### 5.3 PROPORCIONALIDADE EM SENTIDO ESTRITO: BALANÇO ENTRE PRIVACIDADE/AUTONOMIA E DEVER DE PROTEÇÃO

A proporcionalidade, em sentido estrito, exige ponderar se o ganho protetivo esperado compensa a intensidade da restrição imposta aos direitos do adolescente. Em abstrato, a proteção contra a autoagressão e o suicídio tem peso máximo, o que tende a favorecer medidas preventivas, sobretudo quando calibradas para acionar apoio familiar e oferecer recursos de acolhimento. Essa premissa encontra ressonância no ECA Digital ao reconhecer que a proteção da saúde mental integra a expressão do melhor interesse e ao exigir medidas que previnam e mitiguem riscos psicossociais no ambiente digital (BRASIL, 2025). No entanto, a intensidade da interferência também pode ser elevada, pois comunicar a responsáveis a atividade de busca do adolescente toca diretamente à sua privacidade e pode afetar sua autonomia progressiva, inclusive desencorajando a busca por informação ou ajuda em momentos críticos.

Nesse equilíbrio, a medida tende a ser proporcional apenas se acompanhada de salvaguardas que reduzam a intensidade da restrição e aumentem a qualidade do ganho protetivo. O próprio ECA Digital estabelece um limite externo importante para impedir que a proteção se converta em pretexto de vigilância: “A regulamentação não poderá, em nenhuma hipótese, autorizar ou resultar na implantação de mecanismos de vigilância massiva, genérica ou indiscriminada” (BRASIL, 2025). Ainda que o alerta parental não seja, em si, vigilância massiva, o dispositivo funciona como cláusula de contenção: exige que a arquitetura de proteção seja excepcional, focalizada, minimizada e orientada ao cuidado, e não ao monitoramento generalizado de comportamentos informacionais. Por isso, a proporcionalidade estrita, aqui, não se resolve com a afirmação genérica de que “salvar vidas justifica tudo”. Ela exige demonstrar que a medida foi concebida para ser excepcional, calibrada, transparente e orientada à proteção efetiva, com governança e revisão periódica, e com conteúdo





comunicacional minimizado e seguro, de modo que a restrição adicional à privacidade seja a menor possível diante de um risco potencialmente extremo (ALEXY, 2015, p. 118; BRASIL, 2018, 2025).

Em síntese, o teste de proporcionalidade indica que alertas parentais baseados em triagem automatizada podem ser compatíveis com a LGPD e com o ECA Digital quando concebidos como intervenção protetiva calibrada e governada, destinada a riscos graves e acompanhada de transparência, minimização e mecanismos efetivos de revisão e correção. Sem esses elementos, especialmente diante do risco reconhecido de comunicações indevidas e da sensibilidade do tema, a medida tende a aproximar-se de uma interferência excessiva na privacidade e na autonomia progressiva do adolescente, contrariando o melhor interesse e os deveres de proteção previstos no marco brasileiro (BRASIL, 2018, 2025; META, 2026b).

## 6 DIRETRIZES DE CONFORMIDADE E GOVERNANÇA PARA ALERTAS PARENTAIS (PROPOSTA NORMATIVA)

A aplicação do teste de proporcionalidade sugere que alertas parentais podem ser juridicamente defensáveis como medida protetiva quando concebidos e operados sob um modelo de proteção por *design*, com transparência, minimização, segurança e governança contínua. Essa conclusão não transforma o alerta em solução naturalmente legítima; ao contrário, reforça que, no ecossistema brasileiro, a legitimidade da intervenção depende do modo como ela é desenhada, do grau de risco psicossocial que efetivamente enfrenta e das salvaguardas capazes de reduzir falsos positivos, estigmatização e efeitos contraproducentes. Na esfera normativa, as diretrizes a seguir sistematizam parâmetros de conformidade inspirados na LGPD e no Estatuto Digital da Criança e do Adolescente, a partir do caso Meta/Instagram (2026), sem pressupor que a medida corporativa, tal como anunciada, já atenda a todas as exigências possíveis em um cenário regulatório nacional (BRASIL, 2018, 2025; META, 2026b).

### 6.1 TRANSPARÊNCIA SIGNIFICATIVA, AVISO CLARO E PREVISIBILIDADE PARA ADOLESCENTES E RESPONSÁVEIS





O primeiro parâmetro é a transparência em sentido forte, isto é, não limitada a termos de uso genéricos. Alertas parentais pressupõem que o adolescente compreenda, de forma acessível e compatível com seu desenvolvimento, que a supervisão está ativa, quais controles estão em vigor e quais consequências podem ocorrer quando certos comportamentos informacionais se repetem. Em medidas protetivas ancoradas em supervisão parental, a transparência deve operar como condição de legitimidade: a plataforma não apenas pode avisar, mas também precisa construir previsibilidade mínima para que o adolescente não seja surpreendido por comunicações a terceiros sobre aspectos íntimos de sua vida informacional. Em termos de governança, isso inclui explicar o caráter probabilístico do mecanismo, reconhecer o risco de erro e esclarecer que a intervenção não equivale a um diagnóstico, mas a um sinal para buscar apoio, evitando mensagens alarmistas ou punitivas (META, 2026b).

A transparência também deve alcançar os responsáveis. Não basta informar que o recurso existe; é necessário explicitar as escolhas relevantes de *design*, como as condições de acionamento (repetição em curto período), o tipo de conteúdo comunicado e a finalidade protetiva. O objetivo não é revelar segredos industriais, mas permitir a compreensão adequada do que está sendo feito com dados de adolescentes e dos limites adotados para reduzir a intrusão e a supernotificação. Nesse ponto, a própria lógica de proteção integral recomenda uma comunicação centrada em cuidado e orientação, e não em controle disciplinar, pois o alerta só cumpre sua finalidade se vier acompanhado de repertório para conversas sensíveis e de encaminhamento seguro (BRASIL, 2025; META, 2026b).

## 6.2 MINIMIZAÇÃO, REDUÇÃO DE INFERÊNCIAS SENSÍVEIS E SEGURANÇA DA COMUNICAÇÃO

O segundo parâmetro é a minimização, entendida como a limitação da coleta, da retenção, da circulação interna e do conteúdo comunicado. Ainda que o gatilho decorra de buscas associadas à autoagressão, a comunicação ao responsável deve, em regra, evitar detalhamento desnecessário de termos pesquisados, horários e histórico, preferindo uma mensagem focada no essencial: houve um padrão de busca que pode indicar necessidade de apoio e existem recursos de suporte e orientação.

O mesmo raciocínio se aplica ao armazenamento: quando a medida é agregada a





uma função protetiva, a retenção prolongada de *logs* ou a criação de perfis de vulnerabilidade devem ser rigorosamente justificadas, sobretudo porque dados comportamentais podem gerar inferências sobre a saúde mental e, por conseguinte, aproximar-se do regime de cautela reforçada aplicável a dados sensíveis (BRASIL, 2018, 2025).

Para compatibilizar proteção integral e autonomia progressiva, o desenho do mecanismo tende a ser mais proporcional quando incorpora uma lógica de intervenção progressiva, em que o primeiro movimento institucional é fortalecer o suporte e o acolhimento ao próprio adolescente, reservando a comunicação parental imediata como medida calibrada para situações em que sinais mais consistentes indiquem risco psicossocial relevante. Nesse horizonte, é decisivo que a plataforma trate o risco como questão de cuidado e prevenção, e não como gatilho para intensificar a vigilância, evitando que a funcionalidade se converta em rotina de monitoramento generalizado incompatível com as balizas legais de proteção elevada, inclusive no que diz respeito à profilaxia sistêmica do ecossistema de recomendação, à proteção por padrão e à vedação de práticas de perfilamento abusivo (BRASIL, 2025).

A minimização deve vir acompanhada de segurança reforçada da comunicação, pois o alerta, em si, pode expor o adolescente a riscos secundários. A escolha do canal (por exemplo, aplicativos de mensagens) exige avaliação de confidencialidade, de autenticação e do risco de acesso indevido por terceiros, além de orientações claras ao responsável para que não compartilhe o conteúdo do alerta e procure ajuda especializada quando necessário. Em termos de desenho institucional, a proteção por padrão no ambiente digital, aliada a deveres de prevenção e segurança, indica que uma medida protetiva não pode criar uma nova via de exposição do adolescente a constrangimento, violência doméstica, retaliação ou chantagem. Portanto, a segurança da informação e a contenção do conteúdo comunicado não são boas práticas opcionais; são condições materiais para que a intervenção permaneça proporcional (BRASIL, 2018, 2025).

Governança, documentação de riscos, revisão periódica e contestabilidade

O terceiro parâmetro é a governança contínua do mecanismo. Alertas parentais baseados em triagem automatizada devem ser tratados como funcionalidade de alto risco, pois operam sobre um tema sensível e assumem, de partida, a possibilidade de falsos positivos. Por isso, a conformidade não se esgota no lançamento: exige





monitoramento de desempenho, revisão de limiares e auditoria de impactos, inclusive para avaliar se o mecanismo atende à finalidade protetiva ou se produz efeitos adversos previsíveis, como a retração da busca por ajuda ou a migração para canais menos seguros. Esse modelo de responsabilização é coerente com a literatura sobre *accountability* algorítmica, que enfatiza que sistemas que afetam direitos exigem mecanismos institucionais de escrutínio, revisão e correção, e não apenas justificativas abstratas (KROLL et al., 2017).

No ecossistema normativo brasileiro, a governança tende a exigir documentação formal. Medidas que envolvem tratamento de dados de crianças e adolescentes para fins que não sejam estritamente necessários à operação do serviço devem ser acompanhadas de mapeamento de riscos e de relatório de impacto, bem como de monitoramento e avaliação da proteção de dados, disponíveis, sob requisição, da autoridade competente, o que indica um dever de registrar as decisões de *design* e de mitigação adotadas (BRASIL, 2025). Paralelamente, a orientação do ECA Digital de revisão regular de ferramentas de inteligência artificial, com participação de especialistas e possibilidade de desabilitar funcionalidades não essenciais, fornece parâmetro para que alertas baseados em detecção automatizada, especialmente quando conectados a experiências de IA conversacional, passem por ciclos periódicos de validação, calibração e controle de qualidade, com foco em segurança e adequação ao público adolescente (BRASIL, 2025; META, 2026b).

Por fim, a governança proporcional deve incluir contestabilidade. Ainda que o ECA Digital trate expressamente da contestação no procedimento de retirada de conteúdo, sua opção legislativa de informar se a identificação decorreu de análise humana ou automatizada e de garantir recurso sugere uma direção regulatória consistente: medidas relevantes acionadas por automação devem oferecer vias claras de revisão e correção (BRASIL, 2025). No mesmo sentido, a LGPD assegura revisão humana de decisões tomadas unicamente com base em tratamento automatizado que afetem interesses do titular, além de informação sobre critérios e procedimentos, observados os segredos comerciais e industriais. Ainda que haja debate sobre a incidência estrita do art. 20 em um alerta protetivo, a lógica de *due process* informacional recomenda, por prudência jurídica e por aderência ao melhor interesse, a oferta de um canal para contestar alertas indevidos, corrigir configurações e obter esclarecimentos adequados sobre o funcionamento geral do mecanismo, inclusive para reduzir danos decorrentes de falsos positivos (BRASIL, 2018; CORRADINI et al.,





2023).

## 7 CONCLUSÃO

Este artigo examinou o caso Meta/Instagram (2026) como referência contemporânea para discutir a legitimidade de alertas parentais acionados por detecção automatizada de risco de autoagressão nas redes sociais, à luz do ordenamento jurídico brasileiro. Partiu-se de um dado documental delimitado: a Meta anunciou que o Instagram passará a enviar alertas a pais e responsáveis vinculados à supervisão parental quando adolescentes tentarem, em curto prazo, buscar termos relacionados a suicídio e automutilação, com oferta de recursos de apoio e reconhecimento da possibilidade de notificações indevidas em razão de falsos positivos (META, 2026b). O caso foi tratado como estudo de governança algorítmica protetiva, sem pressupor imposição judicial específica, para estabelecer critérios de validade jurídica de medidas semelhantes, nos termos da LGPD e do Estatuto Digital da Criança e do Adolescente.

A análise indicou que a finalidade de prevenir danos graves associados ao suicídio e à automutilação possui elevado peso normativo no sistema de proteção integral, o que torna plausível, em tese, a adoção de intervenções digitais voltadas à mitigação de riscos psicossociais quando calibradas e justificadas. O ECA Digital reforça essa orientação ao exigir medidas razoáveis de prevenção e mitigação desde a concepção e ao longo da operação de produtos e serviços de acesso provável por crianças e adolescentes, além de estabelecer proteção elevada por padrão, instrumentos de supervisão parental e exigências de revisão e de governança de ferramentas de IA e de suporte emocional quando identificados riscos psicossociais (BRASIL, 2025). Em paralelo, a LGPD impõe que o tratamento de dados pessoais de crianças e adolescentes observe o melhor interesse e se submeta a princípios como finalidade, necessidade, transparência, segurança e prevenção, além de oferecer referências relevantes para contestabilidade e revisão quando decisões se apoiam em tratamento automatizado com impacto sobre os interesses do titular (BRASIL, 2018).

Em resposta à pergunta de pesquisa, a aplicação do teste de proporcionalidade indica que alertas parentais baseados em detecção automatizada de risco de autoagressão em redes sociais podem ser considerados proporcionais e compatíveis





com a LGPD e com o ECA Digital apenas quando concebidos como intervenção excepcional e de baixo impacto, sustentada por salvaguardas capazes de reduzir a intensidade da restrição à privacidade e de aumentar a qualidade do ganho protetivo (ALEXY, 2015, p. 116–120). Em termos concretos, isso pressupõe transparência significativa e aviso claro sobre a vigência e o alcance da supervisão, minimização do conteúdo comunicado e da retenção de dados associados ao gatilho, segurança reforçada dos canais de comunicação, governança contínua com revisão periódica dos critérios de acionamento e documentação de riscos e medidas mitigatórias, bem como mecanismos efetivos de contestação e correção diante de alertas indevidos. Na ausência dessas salvaguardas, a medida tende a se aproximar de interferência excessiva, sobretudo porque o próprio desenho reconhece a possibilidade de erros e porque o tema envolve inferências potencialmente sensíveis sobre saúde mental, com risco de efeitos adversos, como estigmatização, escalada de conflitos familiares e retração da busca por ajuda.

Como agenda de pesquisa e de política regulatória, permanece necessário avançar em evidências sobre a efetividade, as taxas de falsos positivos e falsos negativos, e os impactos desiguais em perfis vulneráveis, bem como em diretrizes específicas para intervenções protetivas em interações conversacionais com IA, tema anunciado no próprio caso e que tende a ampliar o espaço de risco e de responsabilidade das plataformas. Em síntese, o caso Meta/Instagram (2026) mostra que a proteção no ambiente digital não se resolve pela escolha entre “privacidade” e “segurança”, mas pelo desenho normativo e técnico de soluções justificáveis, auditáveis e compatíveis com o melhor interesse e com a cidadania digital de crianças e adolescentes no Estado de Direito (BRASIL, 2018, 2025; META, 2026b).

## REFERÊNCIAS

ALEXY, Robert. **Teoria dos Direitos Fundamentais**. Tradução Virgílio Afonso da Silva. 2. ed. São Paulo: Malheiros Editores, 2015.

BÉJAR, Arturo. **Teen Accounts, Broken Promises: HOW INSTAGRAM IS FAILING TO PROTECT MINORS**. [S.l.: s.n.], set. 2025. Disponível em: <https://mollyrosefoundation.org/wp-content/uploads/2025/09/Teen-Accounts-Broken-Promises-How-Instagram-is-failing-to-protect-minors.pdf>. Acesso em: 26 fev. 2026.





BRASIL. **Constituição Federal.** Disponível em: [http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acesso em: 2 fev. 2026.

\_\_\_\_\_. **Lei 8.069/90 (ECA).** Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l8069.htm](https://www.planalto.gov.br/ccivil_03/leis/l8069.htm). Acesso em: 17 fev. 2026.

\_\_\_\_\_. **Lei 13.709/18.** Disponível em: [http://www.planalto.gov.br/ccivil\\_03/ato2015-2018/2018/lei/l13709.htm](http://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/l13709.htm). Acesso em: 7 jan. 2024.

\_\_\_\_\_. **Lei n. 15.211/2025 - Estatuto Digital da Criança e do Adolescente.** Disponível em: [https://www.planalto.gov.br/ccivil\\_03/ato2023-2026/2025/lei/L15211.htm](https://www.planalto.gov.br/ccivil_03/ato2023-2026/2025/lei/L15211.htm). Acesso em: 26 fev. 2026.

\_\_\_\_\_. **Resolução CONANDA n. 257/2024.** Disponível em: <https://www.gov.br/participamaisbrasil/blob/baixar/61597>. Acesso em: 27 fev. 2026.

CANTARINI, Paola. INTELIGÊNCIA ARTIFICIAL, DIREITOS FUNDAMENTAIS POR DESIGN E ESTADO DEMOCRÁTICO DE DIREITO DESDE A CONCEPÇÃO. **Revista Jurídica**, v. 4, n. 76, p. 767–797, mar. 2023. Disponível em: <https://revista.unicuritiba.edu.br/index.php/RevJur/article/view/6681>. Acesso em: 27 fev. 2026.

CORRADINI, Maria Clara Giassetti Medeiros et al. RIGHT TO EXPLANATION: HISTORICAL ANALYSIS OF ART. 20 OF THE LGPD. **Revista Jurídica Unicuritiba**, v. 3, n. 83, p. 347–360, 2025. Disponível em: <https://revista.unicuritiba.edu.br/index.php/RevJur/article/view/8073/371375533>. Acesso em: 26 fev. 2026.

KERR, Dara. **Instagram to alert parents if teens repeatedly search self-harm terms.** Disponível em: <https://www.theguardian.com/technology/2026/feb/26/instagram-meta-self-harm-searches>. Acesso em: 26 fev. 2026.

KROLL, Joshua A et al. Accountable Algorithms. **University of Pennsylvania Law Review**, n. 3, p. 633–705, fev. 2017.

META. **A trajetória da Meta na proteção dos adolescentes e no apoio aos pais.** Disponível em: <https://about.fb.com/br/news/2026/01/a-trajetoria-da-meta-na-protecao-dos-adolescentes-e-no-apoio-aos-pais/>. Acesso em: 27 fev. 2026.

\_\_\_\_\_. **New Alerts to Let Parents Know if Their Teen May Need Support.** Disponível em: <https://about.fb.com/news/2026/02/new-meta-alerts-let-parents-know-if-teen-may-need-support/>. Acesso em: 26 fev. 2026.

REUTERS. **Instagram to alert parents on teen suicide searches as UK weighs social media ban.** Disponível em: <https://www.reuters.com/world/americas/instagram-alert-parents-teen-suicide-searches-uk-weighs-social-media-ban-2026-02-26/>. Acesso em: 27 fev. 2026.





UNICEF. **Guidance on AI and children 3.0.** Disponível em: <https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf>. Acesso em: 27 fev. 2026.

\_\_\_\_\_. **Policy guidance on AI for children: Draft guidance.** Disponível em: <https://criancaeconsumo.org.br/wp-content/uploads/2022/06/unicef-global-insight-policy-guidance-ai-children-20-2021pdf.pdf>. Acesso em: 27 fev. 2026.

UNITED NATIONS. General comment No. 14 (2013) on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1). **Committee on the Rights of the Child**, n. 14, 29 maio 2013. Disponível em: [https://www2.ohchr.org/english/bodies/crc/docs/gc/crc\\_c\\_gc\\_14\\_eng.pdf](https://www2.ohchr.org/english/bodies/crc/docs/gc/crc_c_gc_14_eng.pdf). Acesso em: 27 fev. 2026.

\_\_\_\_\_. General comment No. 25 (2021) on children's rights in relation to the digital environment. **Committee on the Rights of the Child**, 2 mar. 2021. Disponível em: <https://www.unicef.org/bulgaria/en/media/10596/file>. Acesso em: 27 fev. 2026.

