

AGENTES DE INTELIGÊNCIA ARTIFICIAL INCLUSIVOS E PROTETORES: PRIVACIDADE, MELHOR INTERESSE DA CRIANÇA NA PROTEÇÃO DIGITAL DE CRIANÇAS E ADOLESCENTES

INCLUSIVE AND PROTECTIVE ARTIFICIAL INTELLIGENCE AGENTS: PRIVACY, THE BEST INTERESTS OF THE CHILD IN THE DIGITAL PROTECTION OF CHILDREN AND ADOLESCENTS

LUCIANA SABBATINE NEVES

Doutora em Direito Público pela Universidade do Vale do Rio dos Sinos; Mestre em Direitos humanos pela Pontifícia Universidade Católica de São Paulo; Advogada; e-mail: luneves@yahoo.com.

TINA FONSECA

Doutoranda no PPG em *Tecnologias da Inteligência e Design Digital* da Pontifícia Universidade Católica de São Paulo (TIDD); pesquisadora subsidiada pelo CNPq, membro do grupo GEMS e da Rede Saúde Planetária Brasil (IEA-USP); e-mail: ra00341800@pucsp.edu.br e tina.fonseca@educativium.com.

RESUMO

Objetivo: o presente artigo analisa em que medida agentes de inteligência artificial podem ser concebidos como instrumentos inclusivos e protetores, aptos a reforçar a privacidade, a proteção de dados pessoais e as salvaguardas informacionais de crianças e adolescentes, com especial atenção àqueles com deficiência, em conformidade com o princípio do melhor interesse da criança.

Metodologia: adota-se abordagem qualitativa e interdisciplinar, combinando análise jurídico-normativa e reflexão conceitual no campo das Tecnologias da Inteligência e do *Design Digital*. O estudo examina o marco jurídico brasileiro aplicável à proteção de dados, aos direitos de crianças e adolescentes e à inclusão da pessoa com deficiência, em diálogo com a literatura especializada sobre governança algorítmica, acessibilidade por concepção e arquiteturas agênticas.

Resultados: os resultados indicam que a transição de modelos conversacionais para agentes de inteligência artificial amplia as possibilidades de mediação protetiva, permitindo identificar riscos, antecipar exposições danosas e implementar medidas proporcionais de proteção em ambientes digitais. Nesse contexto, a acessibilidade por concepção revela-se requisito central de segurança, na medida em que favorece a compreensão de alertas, contextos de interação e mecanismos de proteção por diferentes perfis de usuários.

Conclusão: ao lado da privacidade por *design* e por padrão, a acessibilidade por concepção ocupa posição estratégica na proteção infantojuvenil digital, oferecendo fundamentos teóricos e jurídicos para o desenvolvimento de tecnologias comprometidas com os direitos humanos, a inclusão e a proteção integral na era digital.

Palavras-chave: Melhor interesse da criança; Agentes de inteligência artificial; Proteção por padrão; Acessibilidade por concepção; Governança algorítmica.

ABSTRACT

Objective: this article analyzes the extent to which artificial intelligence agents may be conceived as inclusive and protective instruments, capable of strengthening the privacy, personal data protection, and informational safeguards of children and adolescents, with special attention to those with disabilities, in accordance with the principle of the best interests of the child.



Methodology: the study adopts a qualitative and interdisciplinary approach, combining legal-normative analysis with conceptual reflection in the field of Intelligence Technologies and Digital Design. It examines the Brazilian legal framework applicable to data protection, the rights of children and adolescents, and the inclusion of persons with disabilities, in dialogue with the specialized literature on algorithmic governance, accessibility by design, and agentic architectures.

Results: the findings indicate that the transition from conversational models to artificial intelligence agents expands the possibilities of protective mediation, enabling the identification of risks, the anticipation of harmful exposure, and the implementation of proportional protective measures in digital environments. In this context, accessibility by design emerges as a central security requirement, insofar as it facilitates the understanding of alerts, interaction contexts, and protection mechanisms by different user profiles.

Conclusion: alongside privacy by design and by default, accessibility by design occupies a strategic position in the digital protection of children and adolescents, providing theoretical and legal foundations for the development of technologies committed to human rights, inclusion, and comprehensive protection in the digital age.

Keywords: Best interests of the child; Artificial intelligence agents; Protection by default; Accessibility by conception; Algorithmic governance.

1 INTRODUÇÃO

A intensificação do uso de sistemas de inteligência artificial em ambientes digitais destinados ou acessados por crianças e adolescentes reconfigura, de maneira profunda, as formas de interação, aprendizagem e cuidado, ao mesmo tempo em que amplia exponencialmente os riscos à privacidade e à proteção de dados pessoais (CGI, 2025). Quando tais sistemas incidem sobre crianças e adolescentes, esses riscos assumem uma dimensão qualitativamente distinta, marcada pela convergência entre vulnerabilidades estruturais e assimetrias informacionais agravadas por limitações de faixa etária, cognitivas, sensoriais ou psicossociais (Neves, 2025).

A crescente incorporação de sistemas de inteligência artificial em ambientes digitais frequentados por crianças e adolescentes, especialmente em plataformas educacionais, jogos *on-line*, redes sociais e aplicações de entretenimento interativo, reconfigura profundamente os processos contemporâneos de socialização, aprendizagem e formação subjetiva (CGI, 2025). Nesse contexto, agentes algorítmicos deixam de operar apenas como ferramentas instrumentais e passam a atuar como mediadores ativos de experiências, escolhas e comportamentos, influenciando diretamente trajetórias de desenvolvimento humano.

Nesse cenário, a tutela jurídica da privacidade não pode ser compreendida de forma isolada ou meramente procedimental, devendo ser articulada com vetores constitucionais, art. 227, Constituição Federal de 1988, como o princípio do melhor interesse da criança e do adolescente, que opera como verdadeiro vetor interpretativo, axiológico e normativo de todo o sistema de proteção infantojuvenil (Pereira, 1999; Tribunal de Justiça do Distrito Federal e dos Territórios, 2025; Frazão, 2021).

No ordenamento jurídico brasileiro, a Constituição Federal consagra a prioridade absoluta da proteção e desenvolvimento integral de crianças e adolescentes, como integrante do núcleo protetivo e formativo da dignidade da pessoa humana, da intimidade, da vida privada e do livre desenvolvimento da personalidade. O Estatuto da Criança e do Adolescente - ECA (Brasil, 1990a), em diálogo com a Lei Geral de Proteção de Dados Pessoais - LGPD (Brasil, 2018), com o Código de Defesa do Consumidor



(Brasil, 1990b) e com a Lei Brasileira de Inclusão da Pessoa com Deficiência (Brasil, 2015), densifica esse comando constitucional ao reconhecer limites rigorosos ao tratamento de dados pessoais, à exploração econômica da atenção infantil e ao uso de tecnologias capazes de influenciar comportamentos, emoções e processos decisórios.

Os avanços normativos associados ao Estatuto Digital da Criança e do Adolescente - Eca Digital, Lei nº 15.211/2025 (Brasil, 2025), reforçam essa arquitetura protetiva, mas a emergência de sistemas de inteligência artificial generativa, preditiva e diagnóstica evidencia a insuficiência de modelos regulatórios centrados exclusivamente na proibição abstrata ou na responsabilização *ex post*, sobretudo quando se trata de crianças e adolescentes, cuja condição demanda proteção reforçada e soluções jurídicas sensíveis à diversidade funcional.

Embora a exigência de revisão humana das decisões automatizadas, já consagrada na LGPD, desempenhe papel relevante nesse contexto, ela se revela claramente insuficiente para assegurar, por si só, a efetividade do melhor interesse da criança. A simples inserção de um *human-in-the-loop* (HITL), ou seja, de um humano que interage ativamente no ciclo de treinamento, supervisão de tomada de decisão de uma IA, não elimina opacidades algorítmicas, não neutraliza vieses estruturais nem resolve dificuldades de comunicação entre sistemas tecnológicos complexos e sujeitos que percebem, interpretam e interagem com o mundo de formas distintas. Torna-se, portanto, necessário avançar para modelos mais densos de governança algorítmica, capazes de incorporar a proteção de direitos fundamentais desde a concepção das tecnologias, *by design* e *by default* (por desenho e por padrão), deslocando o eixo regulatório da reação para a prevenção e para o cuidado, nos termos previstos no Eca Digital.

É a partir dessa premissa que o presente artigo se estrutura em torno da seguinte indagação central: de que modo agentes de inteligência artificial podem ser concebidos e regulados como instrumentos inclusivos e protetores, capazes de assegurar a privacidade e a autodeterminação informacional de crianças e adolescentes, à luz do princípio do melhor interesse da criança?

Como hipótese, sustenta-se que tais agentes não devem ser compreendidos apenas como tecnologias de risco, mas podem ser concebidos como instrumentos jurídicos e tecnológicos de proteção ativa, desde que estruturados segundo princípios de proteção de dados por concepção e por padrão, com mecanismos de explicabilidade adaptados à compreensão de crianças e adolescentes (isto é, capazes de explicar, em linguagem simples e multimodal, por que determinada informação foi bloqueada, por que um alerta foi emitido ou por que um contato foi considerado de risco), além de minimização de dados e medição informacional, operando concretamente como garantidores do melhor interesse da criança com deficiência nos ambientes digitais.

Com base nessa hipótese, o objetivo geral do artigo consiste em analisar em que medida agentes de inteligência artificial podem ser concebidos em arquitetura juridicamente voltada à proteção do melhor interesse de crianças e adolescentes, como instrumentos inclusivos e protetores, aptos a assegurar a privacidade e a autodeterminação informacional de crianças e adolescentes com deficiência e em conformidade com o princípio do melhor interesse da criança, mas não somente.

Para alcançar esse propósito, desenvolve-se, em primeiro lugar, uma análise do conteúdo normativo do referido princípio e de sua incidência específica sobre o desenvolvimento, a



implementação e a operação de agentes de inteligência artificial em ambientes digitais. Em seguida, investigam-se os fundamentos jurídicos da privacidade e da autodeterminação informacional de crianças e adolescentes com deficiência no ordenamento jurídico brasileiro, com especial atenção à Constituição Federal, à LGPD, ao Estatuto da Criança e do Adolescente e à Lei Brasileira de Inclusão. Por fim, identificam-se os limites e as possibilidades da aplicação dos princípios de proteção de dados *por design* e *por padrão* na concepção de agentes de inteligência artificial inclusivos, avaliando-se seu potencial para reduzir riscos, mitigar vulnerabilidades e promover uma arquitetura tecnológica orientada pela proteção integral.

Do ponto de vista metodológico, o estudo adota abordagem qualitativa e interdisciplinar, combinando análise jurídico-normativa e reflexão conceitual no campo das Tecnologias da Inteligência e do Design Digital. Foram examinados dispositivos legais relacionados ao melhor interesse de crianças e adolescentes no ambiente digital, bem como literatura especializada na governança algorítmica, acessibilidade por concepção e arquiteturas agênticas.

Ao articular privacidade, melhor interesse da criança, proteção digital e inclusão, o artigo defende que a regulação da inteligência artificial deve transcender uma lógica meramente reativa e adotar uma perspectiva propositiva e estruturante, na qual agentes de inteligência artificial inclusivos e protetores funcionem como extensões institucionais da proteção integral assegurada pelo Direito. Busca-se, assim, contribuir para o debate nacional e internacional sobre infância, adolescência, inclusão e governança algorítmica, oferecendo fundamentos teóricos e jurídicos para a construção de tecnologias alinhadas aos direitos humanos na era digital.

2 DOS MODELOS CONVERSACIONAIS AOS AGENTES DE IA

Antes de tratar dos agentes, é necessário elucidar o que se entende por inteligência artificial no enquadramento institucional brasileiro: o Plano Brasileiro de Inteligência Artificial – PBI, descreve inteligência artificial como um conjunto de modelos, técnicas e métodos computacionais capazes de produzir classificações, previsões, recomendações e decisões a partir de dados e processos de aprendizagem, com capacidade de influenciar ambientes físicos e digitais (Ministério da Ciência, Tecnologia e Inovação; Centro de Gestão e Estudos Estratégicos, 2025). Esse enquadramento é útil porque impede de reduzir a inteligência artificial a *chatbot* ou tratá-la como entidade antropomorfizada. A inteligência artificial deve, portanto, ser compreendida como tecnologia de decisão sob finalidades e limites definidos por pessoas e instituições.

A passagem de modelos de linguagem ou conversacionais de IA (LLMs), como GPT, Claude, Gemini e Copilot, para modelos de IA de ação larga (LAMs) explica-se pelo fato de o sistema deixar de atuar apenas como produtor de conteúdo e passar a realizar ações verificáveis no ambiente digital. Essa transição muda o centro de gravidade do risco e, conseqüentemente, o centro de gravidade da proteção. Um sistema conversacional pode orientar, advertir e esclarecer; um sistema com agência pode interromper, bloquear, reduzir exposição, registrar ocorrências e acionar uma rede de cuidado.



Essa distinção é coerente com o núcleo técnico clássico do tema: Stuart Russell e Peter Norvig (2021) definem “agente” como a entidade que percebe o ambiente por meio de sensores e age sobre esse ambiente por meio de mecanismos de ação. Além disso, consideram o comportamento do agente como uma função que transforma sequências de percepções em ações. O ponto decisivo, portanto, não é o que o sistema diz, mas como ele decide agir em situações concretas.

Cabe aqui mencionar o projeto WAI – Web Agêntica Inclusiva: protocolos e *proxies* cognitivos para a acessibilidade por concepção, voltado à criação de agentes de IA capazes de mediar a relação entre pessoas com deficiência e perfis neurodivergentes e o ambiente da web, oferecendo suporte embarcado no próprio dispositivo. Considerando, porém, a vulnerabilidade dos demais usuários - ainda que não apresentem deficiência ou incapacidades -, e para além dos agentes destinados à tradução intersemiótica e ao apoio à geração de conteúdo paralinguístico, decidiu-se ampliar essa família agêntica inclusiva com modelos aptos a implementar medidas protetivas, como a identificação e o alerta para sinais de risco, a antecipação de exposições danosas, a reorganização de fluxos excessivamente persuasivos e a oferta de rotas de saída, bloqueios proporcionais e revisão humana em situações especialmente sensíveis, como processos de aliciamento, com vistas à proteção de crianças e adolescentes (Gatti *et al.*, 2026).

A caracterização contemporânea de agentes apresentada pelo Fórum Econômico Mundial (2024) reforça essa mudança de função ao descrever agentes como sistemas capazes de perseguir objetivos, operando em ambientes digitais e físicos com menor dependência da instrução humana contínua, sobretudo quando conectados a modelos de linguagem e a modelos multimodais. Sendo assim, a inteligência artificial deixa de ser apenas uma interface de conversação e passa a ser um componente operacional que integra planejamento, memória e acesso a ferramentas, de modo a concluir tarefas. A diferença é relevante para a proteção infantojuvenil porque desloca o problema do "conteúdo inadequado" para a "trajetória de exposição", isto é, para cadeias de decisões automatizadas que podem conduzir a contato indevido, coleta excessiva de dados, publicidade dirigida e estímulos persistentes ao engajamento.

É nesse contexto que fica evidente a diferença entre modelos centrados na geração de linguagem e arquiteturas centradas em ação, frequentemente denominadas *large action models* (LAM). O que distingue essas arquiteturas não é apenas uma camada de ferramentas, mas a capacidade de transformar intenção em ação dentro de limites estabelecidos. Quando essa capacidade é aplicada ao universo de crianças e adolescentes, a proteção que até aqui se limitava a um simples aconselhamento salta para engenharia preventiva. Isso exige um agente que reconheça sinais de risco e opere intervenções proporcionais: interromper uma tentativa de migração de conversa para canal externo, reduzir exposição à publicidade dirigida, impedir solicitações indevidas de dados, bloquear conteúdos sensíveis, orientar a busca por ajuda e, quando necessário, registrar uma ocorrência mínima para auditoria.

O enquadramento regulatório brasileiro já oferece sinais claros do tipo de requisito que precisa ser convertido em medida técnica. A Autoridade Nacional de Proteção de Dados – ANPD (2025), no que concerne a crianças e adolescentes no ambiente digital, aponta riscos associados a perfilamento, verificação etária e coleta de dados biométricos, além de vincular o tema ao Eca Digital e



à agenda de fiscalização. Em termos práticos, isso significa que a proteção não é compatível com soluções que dependem de coleta ilimitada, com opacidade de critérios ou com intervenções que não podem ser verificadas. Um agente protetor precisa ter finalidade explícita, limites claros e evidências auditáveis de seu funcionamento.

De acordo com Neves (2025), tal transformação impõe desafios inéditos ao Direito, sobretudo no que concerne à proteção da população infantojuvenil, reconhecida constitucionalmente como grupo hipervulnerável (vulnerabilidade agravada). A assimetria estrutural entre o grupo destacado e grandes players dos ecossistemas digitais, marcada por opacidade algorítmica, coleta massiva de dados, arquitetura persuasiva e monetização da atenção, evidencia um cenário de vulnerabilidade ampliada, que demanda respostas normativas capazes de ir além dos modelos tradicionais de consentimento e responsabilização *ex post* (CGI, 2025; Frazão, 2021).

No ordenamento jurídico brasileiro, a promulgação do Estatuto Digital da Criança e do Adolescente, Lei nº 15.211/2025 (Brasil, 2025), em diálogo com a Lei Geral de Proteção de Dados Pessoais, Lei nº 13.709/2018 (Brasil, 2018), inaugura um novo paradigma regulatório fundado na proteção integral, no melhor interesse da criança e na responsabilidade proativa dos agentes tecnológicos, microssistema jurídico que desloca o eixo da tutela de uma lógica reativa para uma lógica preventiva e estrutural, impondo deveres de *privacy by design*, *privacy by default*, verificação etária, mitigação de riscos e vedação à exploração comercial comportamental de dados infantojuvenis.

3 ACESSIBILIDADE POR CONCEPÇÃO

Para crianças e adolescentes com deficiência visual, auditiva e para perfis neurodivergentes, a proteção não pode ser tratada como camada adicional aplicada a um ambiente que permanece hostil. A interface é parte do risco. Quando um usuário não consegue perceber sinais de perigo, compreender avisos, distinguir publicidade de conteúdo, identificar manipulações ou controlar configurações, ele se torna vulnerável por desenho, não por falha individual. Por isso, a *acessibilidade por concepção* é requisito de segurança, tanto quanto de inclusão.

As Diretrizes de Acessibilidade para Conteúdo Web (WCAG), consolidadas pelo *World Wide Web Consortium - W3C* (2023), oferecem critérios testáveis para tornar conteúdos e interfaces mais perceptíveis, operáveis, compreensíveis e robustos, com aprimoramentos relevantes na versão 2.2, inclusive em aspectos de interação e barreiras cognitivas.

Complementarmente, a ABNT NBR 17225 fixa requisitos para acessibilidade em conteúdo e aplicações web no contexto brasileiro, ampliando a base normativa para exigências concretas (ABNT, 2024). É fato que, se o ambiente digital precisa proteger, ele deve ser compreensível; se precisa ser compreensível para todos, deve ser acessível.

Esse eixo permite definir agentes protetores como mediadores comunicacionais e perceptivos. Um agente protetor não se limita a filtrar conteúdo; ele traduz risco para o canal perceptivo e cognitivo do usuário, reduzindo a necessidade de inferência e de leitura implícita. A teoria de Plaza, ao discutir processos de tradução entre sistemas de significação, oferece base para compreender essa



mediação como tradução intersemiótica: sinais de risco precisam ser convertidos em sinais acessíveis ao usuário (Plaza, 2003). Em termos concretos, a criança não pode ser protegida se o aviso de risco permanece em um canal que ela não acessa:

A Tradução Intersemiótica se pauta, então, pelo uso material dos suportes, cujas qualidades e estruturas são os interpretantes dos signos que absorvem, servindo como interfaces. Sendo assim, o operador tradutor, para nós, é mais do que a “interpretação dos signos linguísticos por outros não-linguísticos” (Plaza, 2003 p.67 apud Fonseca. 2012 p.87).

Para crianças surdas e ensurdecidas, parte do risco pode estar no áudio que ela não pode acessar ou mesmo no conteúdo paralinguístico (nas legendas, neste caso): insistência coercitiva, pressão, ameaça, manipulação emocional, ou mesmo conteúdo sonoro inadequado em vídeos e transmissões. Um agente protetor orientado à acessibilidade deve detectar padrões de risco em áudio e convertê-los em aviso visual claro, com linguagem compatível com a faixa etária e, quando aplicável, recursos de língua de sinais em interface. O objetivo não é moralizar interação, mas impedir escaladas de dano e tornar o risco perceptível. A intervenção proporcional inclui avisar, oferecer rota de saída, reduzir exposição e, em situações reiteradas, bloquear continuidade. Concretizando por *design* as diretrizes legais estabelecidas.

No caso das crianças cegas ou com baixa visão, o risco pode ser essencialmente visual: imagens de nudez, violência, botões camuflados, publicidade disfarçada e elementos manipulativos, ou seja, capazes de induzir a geração de imagens mentais enganosas. Um agente protetor, nesse caso, deve agir antes de elas serem expostas a isto, identificando previamente o risco em imagens e vídeos e optando por uma audiodescrição segura ou mesmo o bloqueio do conteúdo, oferecendo uma explicação acessível e a possibilidade de revisão humana em contexto apropriado. O ponto central é evitar que a proteção seja sempre tardia, corretiva.

Quanto a perfis neurodivergentes, a vulnerabilidade pode estar na sobrecarga de estímulos, na imprevisibilidade de interface, em fluxos longos, confusos e em estratégias de persuasão que se apoiam no princípio da escassez, recompensas intermitentes e rolagem infinita. Nesse caso, o agente protetor opera como organizador da experiência, reduzindo distrações, tornando fluxos lineares, exigindo confirmações reforçadas em decisões sensíveis e bloqueando padrões de manipulação recorrentes. O critério é assegurar que escolhas partam de algo que foi, de fato, compreendido.

Esse conjunto de funções estabelece que a acessibilidade - para além de um dever inclusivo - ajude a impedir o dano. Quando a interface é acessível, ela reduz a probabilidade de exploração por engano, pressão ou incompreensão. O ambiente digital, então, deixa de pressupor que a criança, sozinha, é capaz de se defender e passa a incorporar a proteção na sua concepção.

No plano internacional, observa-se a consolidação de uma agenda normativa orientada por inclusão, segurança e direitos humanos, conforme evidenciado por instrumentos como o *Unicef Policy Guidance on AI for Children* (2021), o *UK Age Appropriate Design Code* (2020) e a *Recommendation of the Council on Artificial Intelligence* da Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2019). Contudo, permanece uma lacuna teórica e prática quanto à operacionalização jurídica do conceito de "agentes de inteligência artificial protetores", especialmente



no Sul Global, onde desigualdades digitais, assimetrias informacionais e fragilidades institucionais potencializam os riscos associados à automação decisória.

4 UMA ARQUITETURA PARA O MELHOR INTERESSE

Quando se diz que o melhor interesse deve orientar produtos e serviços digitais acessados por crianças e adolescentes, não se trata de uma cláusula moral abstrata. Trata-se de um vetor normativo de projeto que precisa ser traduzido em arquitetura técnica. A proteção efetiva não nasce de avisos e termos de uso, mas de decisões clássicas da engenharia de software: o que o sistema permite, o que ele impede, quais dados ele coleta, quais inferências ele produz, como ele explica seus atos e como ele reduz o risco de dano previsível. O próprio princípio de privacidade por concepção, formulado por Ann Cavoukian (2009), parte da premissa de que a privacidade deve ser padrão, com minimização de dados, limitação de finalidade e redução de circulação desnecessária de informações pessoais. Para o universo infantojuvenil, essa premissa se torna ainda mais exigente, porque dados e perfis produzidos na infância podem gerar efeitos persistentes e assimétricos no tempo.

Um problema recorrente em soluções protetivas é o paradoxo da proteção que vigia. Muitos sistemas de supervisão dependem de centralização de registros, tais como conteúdos, históricos e padrões de uso que são enviados a servidores externos, formando repositórios sensíveis. Essa estratégia, embora sedutora por oferecer controle, frequentemente aumenta o risco ampliando a superfície de ataque, criando possibilidades de vazamento e abrindo espaço para usos secundários e incompatíveis com a finalidade original. Uma arquitetura alinhada ao melhor interesse tende a privilegiar processamento local sempre que possível, reduzindo a exposição e mantendo a privacidade no dispositivo. A literatura sobre computação na borda (AWS, *s.d.*) descreve exatamente essa migração de processamento para mais perto do usuário, com ganhos de latência e, em muitos casos, de privacidade (Satyanarayanan, 2017). No contexto protetivo, um agente pode detectar sinais de risco e intervir sem exportar dados sensíveis da criança.

Essa escolha arquitetônica tem implicações diretas para agentes protetores. Em vez de registrar conversas inteiras, imagens ou históricos completos, o agente pode produzir registros mínimos e tecnicamente úteis: ocorrência de tentativa de contato não verificada, insistência repetida, presença de padrão de engano, solicitação de dado sensível, exposição a conteúdo inadequado, tentativa de transação. Esse tipo de registro não substitui investigação humana, mas permite auditoria e responsabilização sem exposição indevida. O critério é simples: registrar a evidência mínima necessária para demonstrar que a proteção foi acionada e para permitir revisão qualificada, sem transformar a vida da criança em objeto de monitoramento contínuo.

O mesmo raciocínio se aplica à verificação etária e a medidas de acesso. O melhor interesse exige evitar duas distorções: a permissividade total, que expõe; e o controle invasivo, que viola privacidade. Nesse ponto, a governança técnica precisa ser desenhada para provar o necessário e ocultar o excedente, evitando a lógica de identificação permanente. A literatura e os padrões técnicos oferecem caminhos para isso, desde que aplicados com rigor e finalidade protetiva.



5 POR UMA WEB AGÊNTICA, PROTETORA E INCLUSIVA

A reorientação protetiva ganha robustez quando é alinhada a compromissos institucionais explícitos. O PBIA adota - como orientação de política pública - a inteligência artificial voltada ao bem-estar social, com princípios éticos e não discriminatórios (MCTI; CGEE, 2025). Essa orientação reforça que agência técnica não é valor em si, portanto deve ser subordinada a finalidades legítimas, entre as quais a proteção integral de crianças e adolescentes ocupa posição prioritária.

No mesmo sentido, a Secretaria de Comunicação Social da Presidência da República - Secom/PR (2025), informou, que o Governo Federal havia criado o Comitê Intersectorial para a Proteção dos Direitos da Criança e do Adolescente no Ambiente Digital. Esse comitê reúne órgãos do governo para propor e coordenar medidas de proteção a crianças e adolescentes na internet. Na mesma publicação, a Secom/PR divulgou a “escuta” nacional sobre uso de telas, feita com o Instituto Alana, e indicou o relatório completo em com os resultados e recomendações.

A contribuição do Instituto Alana (*s.d.*), reconhecida no processo de escuta, também fortalece o eixo normativo de que a internet deve ser organizada segundo o melhor interesse, com especial atenção à proteção contra exploração comercial infantil e contra formas de persuasão voltadas a vulnerabilidades. Isso se conecta diretamente à crítica contemporânea à economia da atenção e à arquitetura persuasiva, que é quando plataformas são desenhadas para maximizar permanência e coleta de dados, operando contra a proteção por concepção. Deste modo, entende-se que o mesmo potencial de agência que hoje sustenta estratégias de engajamento pode, por outro desenho de finalidade e de limites, sustentar proteção.

No plano internacional, diretrizes da Unicef (2021) sobre inteligência artificial e crianças apontam requisitos para que os sistemas respeitem direitos e reduzam riscos: segurança, privacidade, transparência e governança, com atenção especial ao fato de que crianças são afetadas por sistemas automatizados mesmo quando não os percebem. Esse ponto é compatível com a tese de que uma proteção efetiva não pode depender de consciência situacional plena da criança, sobretudo quando se trata de um risco invisível e estrutural. Por isso, agentes protetores são necessários: por perceberem o ambiente, aplicarem regras de segurança e intervirem de forma proporcional, acessível e verificável.

Entendemos que quando sistemas deixam de ser instrumentos e passam a mediar condutas, a proteção exige arquitetura e governança. Arquitetura para o melhor interesse, acessibilidade por concepção e governança verificável são condições elementares para que a agência técnica opere em favor de crianças e adolescentes, especialmente daqueles com deficiência e/ou perfis neurodivergentes. Trata-se de construir ambientes digitais nos quais a agência técnica esteja subordinada a direitos e a internet opere em favor da proteção e do bem-estar.



6 GOVERNANÇA ALGORÍTMICA E IDENTIDADE VERIFICÁVEL

Sistemas que percebem e agem sobre o ambiente digital precisam ser desenhados para reduzir dano previsível, registrar o essencial para auditoria e permitir revisão humana qualificada. Shneiderman (2022), ao defender uma inteligência artificial centrada no humano, insiste em sistemas confiáveis e seguros, com validação, verificação, trilhas de auditoria, testes e supervisão que conectem princípios éticos a procedimentos técnicos. Essa orientação é particularmente importante em proteção infantojuvenil porque as consequências de erro são graves e, muitas vezes, irreversíveis. Assim, a governança deve estar baseada em evidências.

Um eixo central dessa governança é a identidade. Riscos como aproximação indevida, falsidade de perfis e escalada de contato se aproveitam de assimetria: a criança não tem meios de verificar quem está do outro lado, e não deve ser responsabilizada por isso. Em arquitetura protetiva, parte desse fardo precisa ser deslocada para a infraestrutura. É nesse ponto que padrões do W3C (s.d.) sobre credenciais verificáveis ganham relevância prática: o modelo de credenciais verificáveis permite provar atributos de forma criptograficamente verificável, com foco em segurança e preservação de privacidade. Isso possibilita construir meios de verificação que não exigem exposição total de identidade e que evitam coleta excessiva.

Aplicado à proteção, esse padrão permite um agente guardião de identidade com função claramente limitada: exigir prova verificável quando a interação envolve risco. Exemplos: um adulto que pretende entrar em canal de conversa com crianças deve provar vínculo institucional ou autorização, sem revelar dados além do necessário; um ambiente escolar pode exigir credenciais verificáveis de pertencimento para habilitar contato entre estudantes; canais privados podem ser bloqueados quando não há prova mínima. O objetivo não é vigiar, mas impedir aproximações indevidas por padrão. A responsabilidade de detectar o perigo deixa, efetivamente, de ser dos usuários (que, vale lembrar, são crianças e adolescentes) e passa a ser do ambiente digital já capaz de coibir oportunidades de exploração.

Para completar a governança, é necessário definir responsabilidade institucional e procedimentos: Quem configura políticas do agente, quem revisa registros mínimos, quem responde a incidentes e como evitar que medidas protetivas se tornem medidas invasivas. A resposta técnica mais consistente combina limites de permissão, registro mínimo e revisão humana qualificada. Isso permite compatibilizar proteção com privacidade. O agente não precisa registrar toda a jornada digital do usuário, apenas o essencial para demonstrar que atuou conforme o melhor interesse e intervir quando necessário.

7 RESULTADOS ESPERADOS

O estudo vislumbra a proteção de crianças e adolescentes incorporada à própria arquitetura dos sistemas digitais. A passagem de modelos conversacionais para agentes de inteligência artificial abre caminho para formas de mediação técnica capazes de reconhecer situações de risco, apoiar



decisões mais seguras e reduzir exposições potencialmente danosas. Nesse contexto, a acessibilidade por concepção assume papel central, pois amplia a inteligibilidade das interfaces e permite que diferentes perfis de usuários compreendam alertas, informações e contextos de interação. Ao articular princípios jurídicos de proteção com recursos da *web agêntica*, este trabalho dispõe de fundamentos para o desenvolvimento de agentes de IA inclusivos e protetores, alinhados ao melhor interesse da criança, à proteção por padrão e à governança algorítmica.

8 CONSIDERAÇÕES FINAIS

A mediação digital passou a ocupar posição estruturante na vida social, educacional e informacional de crianças e adolescentes. Sistemas inteligentes participam da organização das interações, da circulação de conteúdos e da formação de escolhas. Esse cenário exige que o dever de cuidado alcance também o desenho das tecnologias que organizam essas experiências. Ao aproximar Direito Público, Tecnologias da Inteligência e Design Digital, este estudo sustenta que agentes de inteligência artificial podem atuar como mediadores técnicos orientados por critérios de proteção e inclusão. A incorporação da acessibilidade por concepção amplia a capacidade de compreensão, fortalece a segurança informacional e contribui para uma web que acolha a diversidade de usuários. O desenvolvimento de agentes inclusivos e protetores aponta, assim, para uma agenda de inovação tecnológica comprometida com a dignidade e a proteção de crianças e adolescentes.

REFERÊNCIAS

ABNT. NBR 17225: **Acessibilidade em conteúdo e aplicações web**: Requisitos. Rio de Janeiro: ABNT, 2024. Disponível em: <https://www.abntcatalogo.com.br>. Acesso em: 3 jan. 2026.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Semana da Criança**: compromisso da ANPD com a proteção de crianças e adolescentes no ambiente digital. 2025. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/semana-da-crianca-compromisso-da-anpd-com-a-protecao-de-criancas-e-adolescentes-no-ambiente-digital>. Acesso em: 29 jan. 2026.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 15 out. 2023.

BRASIL. **Lei nº 8.069, de 13 de julho de 1990**. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências. Diário Oficial da União: seção 1, Brasília, DF, 16 jul. 1990. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8069.htm. Acesso em: 15 out. 2023.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências (Código de Defesa do Consumidor). Diário Oficial da União: seção 1, Brasília, DF, 12 set. 1990. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8078.htm. Acesso em: 3 fev. 2026.



BRASIL. **Lei nº 13.146, de 6 de julho de 2015**. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência). Diário Oficial da União: seção 1, Brasília, DF, 7 jul. 2015. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 3 fev. 2026.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 15 out. 2023.

BRASIL. **Lei nº 15.211, de 17 de setembro de 2025**. Institui o Estatuto Digital da Criança e do Adolescente. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/L15211.htm. Acesso em: 22 dez. 2025.

CAVOUKIAN, Ann. **Privacy by Design: The 7 Foundational Principles**. 2009. Disponível em: <https://www.ipc.on.ca/wp-content/uploads/2013/09/pbd-primer.pdf>. Acesso em: 3 jan. 2026.

COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.br). **Pesquisa TIC Kids Online Brasil**: uso da Internet por crianças e adolescentes no Brasil. São Paulo: NIC.br; Cetic.br, 2025. Disponível em: <https://cetic.br/pt/pesquisa/kids-online/>. Acesso em: 22 dez. 2025.

FRAZÃO, Ana. **Dever geral de cuidado das plataformas diante de crianças e adolescentes**. São Paulo: Instituto Alana, 2021.

GATTI, Daniel Couto *et al.* **WAI - Web Agêntica Inclusiva**: protocolos e proxies cognitivos para a acessibilidade por concepção. São Paulo: Pontifícia Universidade Católica de São Paulo, 2026. Projeto de pesquisa em desenvolvimento.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Age appropriate design: a code of practice for online services**. Wilmslow: ICO, 2020. Disponível em: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/>. Acesso em: 14 fev. 2026.

INSTITUTO ALANA. **Criança e Consumo**. Disponível em: <https://alana.org.br/instituto-alana/iniciativas/crianca-e-consumo>. Acesso em: 14 fev. 2026.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI); CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE). **Plano Brasileiro de Inteligência Artificial (PBIA)**: IA para o bem de todos. Brasília, DF: MCTI; CGEE, 2025. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/plano-brasileiro-de-inteligencia-artificial-pbia- vf.pdf>. Acesso em: 27 fev. 2026.

NEVES, Luciana Sabbatine. **Metaverso e consumidores hipervulneráveis**: fundamentos para a positivação dos direitos à privacidade, integridade cognitiva e do sistema nervoso de crianças e adolescentes, no Brasil e no MERCOSUL. 2025. Tese (Doutorado em Direito) – Universidade do Vale do Rio dos Sinos, São Leopoldo, 2025. Disponível em: <https://www.repositorio.jesuita.org.br/handle/UNISINOS/1646>. Acesso em: 29 jan. 2026.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Recommendation of the Council on Artificial Intelligence**. Paris: OCDE, 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em: 14 fev. 2026.

PEREIRA, Tânia da Silva. **O “melhor interesse da criança”**. In: PEREIRA, Tânia da Silva (coord.). O melhor interesse da criança: um debate interdisciplinar. Rio de Janeiro: Renovar, 1999. p. 3-.



PLAZA, Júlio. **Tradução intersemiótica**. São Paulo: Perspectiva, 2003.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 4. ed. 2021. Disponível em: <http://aima.cs.berkeley.edu/>. Acesso em: 21 dez. 2025.

SATYANARAYANAN, Mahadev. **The emergence of edge computing**. *Computer*, v. 50, n. 1, p. 30–39, 2017. Disponível em: <https://ieeexplore.ieee.org/document/7807168>. Acesso em: 12 jan. 2026.

SECRETARIA DE COMUNICAÇÃO SOCIAL DA PRESIDÊNCIA DA REPÚBLICA (SECOM). **Governo anuncia comitê interministerial para a proteção de crianças e adolescentes no ambiente digital**. 2025. Disponível em: <https://www.gov.br/secom/pt-br/acompanhe-a-secom/noticias/2025/07/governo-anuncia-comite-interministerial-para-a-protecao-de-criancas-e-adolescentes-no-ambiente-digital>. Acesso em: 14 fev. 2026.

SHNEIDERMAN, Ben. **Human-Centered AI**. Oxford: Oxford University Press, 2022. Disponível em: <https://academic.oup.com/book/41126/chapter-abstract/350464263>. Acesso em: 20 fev. 2026.

TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E DOS TERRITÓRIOS (TJDFT). **Princípio da proteção integral**: direito da criança e do adolescente. Brasília, DF, 23 jul. 2025. Disponível em: <https://www.tjdft.jus.br/consultas/jurisprudencia/jurisprudencia-em-temas/direito-da-crianca-e-do-adolescente-na-visao-do-tjdft/dos-principios-e-fundamentos/principio-da-protecao-integral-direito-das-criancas-e-dos-adolescentes>. Acesso em: 29 jan. 2026.

UNICEF. **Policy guidance on AI for children**. 2021. Disponível em: <https://www.unicef.org/innocenti/reports/policy-guidance-ai-children>. Acesso em: 14 fev. 2026.

WORLD ECONOMIC FORUM. **Navigating the AI Frontier: A Primer on the Evolution and Impact of AI Agents**. 2024. Disponível em: https://www3.weforum.org/docs/WEF_Navigating_the_AI_Frontier_AI_Agents_2024.pdf. Acesso em: 29 jan. 2026.

WORLD WIDE WEB CONSORTIUM (W3C). **Web Content Accessibility Guidelines (WCAG) 2.2**. 2023. Disponível em: <https://www.w3.org/TR/WCAG22/>. Acesso em: 21 dez. 2025.

Declaração de uso de inteligência artificial

Foi utilizada ferramenta de inteligência artificial como apoio à revisão linguística, correção textual e organização do manuscrito. A ferramenta também foi empregada para apoio à verificação preliminar de fontes públicas. Todas as informações foram revisadas e validadas pelas autoras, que assumem integral responsabilidade pelo conteúdo final do texto.

