JCOTS

JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Supplementary Report

AI Chatbots: Companionship & Minors

August 2026

How to read this report

This report is a supplement to the 2025 JCOTS report on AI Chatbots. It is organized into eight sections:

- The **executive summary** provides an overview of the report's key themes and recommendations;
- The **study background** explains the impetus and scope of the report;
- **Key concepts** provides definitions for important or specialized terms used throughout the report;
- The **introduction** provides an overview of the policy issues around AI chatbots and emotional and relational harms;
- **Revisiting AI chatbots** recaps several key points from the previous 2025 JCOTS study on AI chatbots and delves more deeply into several topics specific to the proposed Virginia legislation, including emotional and relational harms and risks, AI companionship, minors' use of chatbots, and lessons to be learned from policy on social media;
- **Virginia proposed legislation** provides an overview of the two chatbot bills introduced in the last General Assembly session;
- **Policy landscape** offers a review of how other states have legislated on chatbots in the last year, congressional proposals at the federal level, and litigation concerning chatbots;
- **Recommendations** sets out several overarching policy principles and policy options for addressing the risks and harms associated with AI chatbots if legislation were to be re-introduced in Virginia.

The report can be read as a continuous document, or readers can skip to sections of interest.

Executive Summary

This report examines the opportunities, risks, and harms associated with using AI chatbots for companionship. It further explores how minors are using AI chatbots and specific policy issues for youth users of these products. This year legislatures across the United States turned their attention to chatbots, introducing around 100 chatbot bills, with 16 signed into law. Legislative action has responded to growing chatbot use; a 2026 Pew survey found that around 50% of American adults and 64% of American teens use chatbots, and many users report turning to chatbots for emotional or mental health support. Alongside this trend, journalism and academic research have begun surfacing concerning risks and harms associated with AI chatbot use, especially sustained, longer-term engagement with chatbots, including cases of delusional spirals, emotional dependence, and suicide. These emotional and relational harms are the result of the complex interactions between humans and machines, which impact people's psychological states, emotional conditions, and social lives.

Beyond the emerging risks of chatbots more broadly, several high-profile tragedies have drawn particular attention to the risks of minors using chatbots. The unique developmental needs of minors suggest that they may be more susceptible to emotional and relational risks. However, the report finds both adults and minors experiencing documented emotional and relational harms correlated with chatbots, and scientific evidence on minors' engagement with chatbots is limited. Documented cases of emotional and relational harms for all users have generally involved turning to chatbots for intimate or romantic interaction or mental health support—types of engagement that might be considered forms of "companionship." While some AI chatbots are specifically marketed as "AI companions" to provide emotional, romantic, or even therapeutic support, the report finds that it is not meaningful from a policy standpoint to distinguish "AI companions" from AI chatbots that happen to be used for companionship. The risks and harms associated with AI chatbots are a product of their underlying technology (Large-Language Models) and design features, regardless of whether they might be categorized as general purpose AI chatbots (e.g. ChatGPT, Gemini, Grok), task- or domain-specific AI chatbots (e.g. therapy bots or educational tutors), or AI models marketed for social purposes (e.g. Character.AI, Replika).

The report identifies several design features associated with emotional and relational risks, including sycophancy, anthropomorphism-enhancing design, conversation prolongation, and weak disengagement mechanisms. There is currently little scientific evidence supporting generalizable, population-level, causal connections between design features and emotional and relational harms due to the relative newness of the technology and the complexity of technological, social, psychological, and environmental factors involved. However, a growing body of evidence suggests AI chatbot use is

correlated with emotional and relational harms, and an emerging scholarly consensus points to engagement-maximizing design as a policy concern. Another policy concern centers on the impediments to learning more about AI chatbot risks and harms: there is currently a lack of independent, verifiable safety testing, risk assessment, incident reporting, or effective safeguards for commercial AI chatbots. Policy on AI chatbots must therefore weigh the risks of legislative inaction to mitigate observed societal harms against the risks of legislative action with untested policy solutions as the evidence base develops and matures.

When it comes to weighing these risks, the report draws a comparison to social media policy. Despite growing public concern and accumulating evidence of individual harms, policy has lagged behind the evolution and market dominance of these platforms. Three relevant lessons can be drawn from how policy has confronted social media: (1) that voluntary commitments by companies have been largely ineffective at addressing harms, (2) scientific evidence alone cannot offer timely and effective policy solutions for the technological harms people are experiencing, (3) a lack of public information and data has delayed interventions. Collectively, these lessons point to the importance of legislative action that enforces harm prevention and simultaneously closes the information gap between companies, the public, and independent researchers.

The report ultimately concludes that although the scientific evidence on these issues is still developing, the observable risks warrant legislative action. First, the extreme harms experienced by some users require mitigation, and AI companies have not demonstrated sufficient commitment to confirming safety on a voluntary basis. Second, more information is needed in order to build independent scientific evidence on how chatbots affect individuals and society, to make AI products safer, and to inform more effective policy. Legislation can play an important role in mandating transparency to enable this public oversight. Therefore, the report recommends re-introducing legislation on AI chatbots that follows six policy principles: (1) mitigate extreme harms for all users; (2) target design features rather than content; (3) mandate transparency and accountability to build the evidence base and facilitate independent, public oversight; (4) enable user choice and autonomy by establishing privacy-respecting defaults and options that accommodate users' preferences; (5) enable legislative learning by reviewing the impact of legislation once enacted; (6) engage affected communities. Within each of these principles, the report presents several policy options for legislation.

Study Background

During the 2026 General Assembly session, JCOTS was sent HB 635 ("Artificial Intelligence Chatbots Act") and SB796 ("Artificial Intelligence Companion Chatbots and Minors Act") for study by letter. These proposed bills would have regulated AI chatbots in the Commonwealth, with particular emphasis on preventing covered harms, such as self-harm or suicide, and requiring additional guardrails for minors.

Over 30 states have introduced almost 100 bills targeting chatbots this year, and as of the end of July, 16 states have passed chatbot legislation. Recent chatbot legislation has overwhelmingly focused on mitigating the emotional and relational risks to users of chatbot products, such as dependence, self-harm, and delusional thinking, and 10 bills passed in other states contained specific provisions for minors. This legislative activity has occurred in spite of the White House Executive Order that discouraged states from passing onerous AI legislation.[1] However, several states focused their regulatory attention on minors due to an exemption articulated in the Executive Order for "child safety protections."

In 2025, JCOTS published an informational report on AI chatbots, which covered a technical overview of LLM-based chatbots, including AI assistants, AI agents, and AI companions, and emergent policy issues associated with these products. Following the two proposed Virginia bills on chatbots and the flurry of state legislative activity on chatbots, this report builds on last year's study, focusing on:

- Chatbot companionship;
- The emotional, social, and psychological risks of chatbots;
- Minors' use of chatbots;
- Chatbot legislation in other states and at the federal level.

Methodology

This report is based on a literature review of academic and grey literature and stakeholder interviews. Generative AI was used for notetaking during some of the interviews and in the process of conducting background research for this study. AI was not used for outlining, writing, or editing this report.

Key Concepts

Age-appropriate design: A set of standards and explicit design requirements for online services to accommodate the developmental needs and protect the safety and privacy of children and adolescents, prioritizing safety defaults and aligning design features with developmental stages.

Age assurance: Systems that attempt to ascertain with varying degrees of precision whether a user fits within a specific, eligible age range to access restricted content or services. Service providers use various methods to conduct this verification, such as age attestation (e.g. inputting a birthdate), age estimation or inference (e.g. algorithmic inference based on user behavior on a platform), or age verification (e.g. uploading a Driver's License).

Companion chatbot: A type of AI designed and marketed specifically to offer emotional support or social interaction, often using natural language processing, sentiment analysis, and personalized behavior to create a more human-like, engaging experience and provide users with a sense of connection and support.

Dark patterns: Deceptive and manipulative design features built into digital platforms that prioritize user engagement over user well-being and are not in the users' best interests.

Emotional and relational harms: Negative outcomes arising from complex interactions between humans and AI chatbots that impact people's psychological states, emotional conditions, and social lives.

Emulated empathy: The process by which an AI model mimics human empathy by statistically predicting and generating language designed to evoke an emotional response, despite the model lacking actual human understanding or experience.

Generative AI (GenAI): AI models that are able to generate new content. When prompted (often by a user inputting text in natural language), GenAI can produce various outputs, including text, images, videos, computer code, or music.

General-purpose chatbot: An AI system primarily designed to support a broad array of functional, communicative, and assistive tasks. In contrast to companion chatbots, general purpose chatbots (e.g. ChatGPT, Claude, Gemini) are task-oriented, although users may also seek emotional or mental health support from them.

Information asymmetry: The imbalance of information between the private and public sectors that occurs when technology companies closely guard data, testing reports, and product access, forcing the government and independent researchers to gather their own data slowly (and often indirectly) to inform public policy.

Large Language Model (LLM): A type of AI that uses deep learning techniques to process and generate human-like text based on vast amounts of data that include books, websites, and other text sources, enabling them to learn patterns, structures, and nuances of language.

Pacing problem: A term from regulatory theory that describes when rapid technological innovation and new product releases outpace both government regulation and the development of rigorous scientific evidence.

Prompt: The input or query provided to an LLM to guide its response, which can be a question, statement, or instruction that sets the context for what the model should generate.

Training data: The huge amount of text data used to teach an LLM how to understand and generate human language, typically consisting of diverse text sources such as books, articles, websites, and other publicly available written content that helps the LLM learn patterns, structures, grammar, vocabulary, and nuances of language by analyzing the relationships between words, phrases, and sentences.

Introduction

In the last several years, AI chatbots have become widespread and constitute a primary avenue through which members of the public engage with artificial intelligence (AI) in daily life. Chatbots are AI systems that users can interact with using natural language rather than computer code. They can perform complex tasks like searching the Internet, summarizing or editing documents, writing code, or generating new content, like reports, videos, and images. Examples of popular AI chatbots include

ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic), CoPilot (Microsoft), Perplexity (Perplexity AI, Inc.), and Grok (SpaceXAI). Social media companies including Meta, Snapchat, and TikTok also have AI chatbots. In addition, there are chatbots marketed for specific purposes, such as tutoring, coding, or romantic relationships.

A 2026 Pew Survey found that around 50% of American adults use chatbots, with 25% using them almost daily.[2] An earlier 2025 Pew Survey found that 64% of American teens use chatbots, with 28% using them almost daily.[3] At the same time, public opinion has increasingly turned against AI, with 40% of adults expressing an expectation that AI will have a negative impact on society in the next 20 years. Although young people are more likely to take a positive view of the impact of chatbots on society (31%), 26% of teens expect chatbots to have a negative impact, and 34% say they will have equally positive and negative effects. As more people of all ages have begun using these products, some users have experienced serious harms as a result of their interactions with chatbots, including delusional spirals, self-harm, suicide, and sexual exploitation and manipulation. Several high-profile tragedies claimed headlines, in which young people died by suicide following destructive conversations with chatbots. Within just the last year, the *New York Times* documented over 50 cases of psychological crisis associated with chatbot use.[4] As of November 2025 OpenAI, the company behind ChatGPT, faced five wrongful death lawsuits.[5]

Chatbots in the News

- "Can AI Be Blamed for a Teen's Suicide?" *The New York Times*, 4 April 2024
- "AI friendships claim to cure loneliness. Some are ending in suicide." *The Washington Post*, 6 December 2024
- " 'This was trauma by simulation.' ChatGPT users file disturbing mental health complaints." *Gizmodo*, 13 August 2025
- "What My Daughter Told ChatGPT Before She Took Her Life," *The New York Times*, 18 August 2025
- "A Teen Was Suicidal. ChatGPT was the Friend He Confided In," *The New York Times*, 26 August 2025
- "I wanted ChatGPT to help me. So why did it advise me how to kill myself?" *BBC*, 6

November 2025

- " 'Our Bond Is the Only Thing That's Real:' A New Lawsuit Alleges Google Gemini Drove a Man to Suicide." *TIME*, 4 March 2026

An Ipsos survey this year reported that 60% of Americans favor government regulation to ensure AI outputs do not cause harm.[6] State legislatures have responded to public concern with a burst of legislation targeting chatbots during 2026 sessions. Nearly 100 bills have been proposed this year, and to date 16 states have passed chatbot regulations. Legislation has predominantly focused on mitigating the social and emotional harms consumers are experiencing, as evidence shows that people are turning to chatbots for emotional, romantic, social, and mental health support. The recent Pew survey found that 14% of adults use chatbots for companionship or emotional support,[7] and a report by the UK's AI Security Institute found that around a third of respondents had used AI chatbots for emotional support in the last year.[8] Other researchers conducting a longitudinal social listening study on chatbot use observed that therapy/companionship is the fastest growing use of chatbots and constitutes the top use of chatbots for the last two years.[9] Research on the psychological and emotional effects of chatbots is at an early stage, but parents, educators, and clinicians have raised alarms based on their lived experiences. A survey conducted by the American Psychological Association found that 39% of psychologists have patients who used AI to self-diagnose, and 35% reported that patients have turned to chatbots to act as an additional mental health professional.[10]

In this context, this report builds on a 2025 JCOTS study on AI chatbots, with a renewed focus on chatbots used for emotional, psychological, or relational support and the related emotional harms that policy may seek to mitigate. This focus encompasses a wide range of chatbots but a narrower set of risks and harms than the previous study. Many state bills introduced this year specifically targeted companion chatbots, as these are marketed for social, emotional, and romantic uses. However, this report takes a broader view to include chatbots capable of engaging in emotionally charged conversations, even if they are marketed for other purposes. As the report will go on to explain below, there is ample evidence that any chatbot with these capabilities may generate the harms targeted in the proposed Virginia legislation that prompted this study.

Revisiting AI Chatbots

AI chatbots have become a popular technology product following OpenAI's seismic release of ChatGPT in 2022. Researchers and technology companies had been working on the underlying

technology, Large Language Models (LLMs), for decades, and a commercial market for powerful LLMs was already developing in the technology sector.[11] However, ChatGPT was the first successful LLM product released to the public. Since then, chatbots marketed to the public have proliferated, and the companies behind them have gained notoriety along with unprecedented valuations and investment. Several anticipated artificial intelligence IPOs this year are set to break records.[12]

A brief overview of the technology behind chatbots is warranted because the technical architecture of chatbots plays an important role in shaping how people engage with and are affected by chatbots. LLMs are a type of AI that uses deep learning techniques to process and generate text outputs in natural human language. They are trained on vast amounts of data that include books, websites, and other text sources, enabling them to identify and replicate the patterns, structures, and nuances of language. LLMs are a subtype of generative AI (sometimes called "genAI"), which refers to any models trained on large volumes of data that are able to generate new content, such as text, images, audio, or video.

Generative AI training practices have come under scrutiny because companies have acquired the vast amounts of data needed for training by scraping as much of the internet as possible, without regard for copyright, intellectual property, or privacy.[13] The training process results in human-sounding outputs that also appear informed and knowledgeable. However, although these data aggregation practices have resulted in models ingesting a huge amount of information, it is not accurate to say that even the largest models are trained on all of human knowledge. Only a small portion of human knowledge exists in written form, and only a fraction of that repository has been digitized.[14] Additionally, only around 7% of the world's languages are well-represented on the internet.[15] Analyses of crowdsourced sites like Wikipedia, where members of the public can contribute information for free, show that content is disproportionately produced by men.[16] Studies of popular AI training datasets also show that they contain large quantities of abusive, sexually explicit, discriminatory, and personally identifiable content.[17] Training data is only one factor in how models are built, but it is the formative, raw material from which models have "learned" how to generate human-like responses. The omissions and ethical ambiguities embedded in training datasets illustrate the complexities of AI safety, even at the most basic layer of model development.

The distinguishing feature of AI chatbots is their ability to respond to natural human language with novel responses that mimic human communication. LLMs do this using complex mathematical operations, which enable the model to predict likely outputs. In other words, although the outputs of LLMs are delivered in the form of human-like language, they are the product of statistics. They are highly capable of mimicking human language, but they lack human understanding and experience.[18]

The probabilistic nature of the outputs means that chatbots generate unpredictable responses, and even model developers do not know in advance how a chatbot will answer a user prompt. This presents many challenges for addressing problematic chatbot outputs because the science is still evolving on how to explain or reverse-engineer a chatbot response. Developers set extensive parameters for models (using model weights, fine-tuning, reinforcement learning, and other methods, including hard-coding guardrails to constrain model outputs), but LLMs still suffer from a "black box" problem: it is often hard, if not impossible, to know with certainty how a model generates a specific output.[19]

A baseline understanding of the technology behind chatbots is helpful in making sense of the policy issues these products raise. Scientific research on LLM performance "in the wild" is an evolving field, but emerging scholarship (in controlled and real-life contexts) has demonstrated that even well-trained models can be prompted to misbehave,[20] well-constrained models can leak important information,[21] and developers themselves cannot reliably reverse-engineer LLM decisions or reasoning.[22] Last year's JCOTS report highlighted several policy areas deserving attention with regard to chatbots in light of these observable risks: health and wellbeing, privacy, security, transparency and explainability, workforce development, and democratic culture. Since then, public interest has focused predominantly on issues in the health and wellbeing category. These issues are what psychologist Marlynn Wei calls "emotional and relational risks." They include a deterioration of boundaries, emotional dependence, emotional manipulation, social isolation, and even psychosis.[23] Emotional and relational risks of chatbots are rooted in the technical architecture of the AI models underlying chatbots. LLMs are probabilistic prediction systems that emulate empathy by statistically generating language that evokes an emotional response in humans, who are hard-wired to interpret social cues like language.[24] Microsoft's AI CEO recently described this effect as "hack[ing] our human empathy,"[25] and others have called it "attachment hacking."[26]

Importantly, it is not always possible to draw a bright line between different categories of chatbots, and the marketing for a chatbot may not accurately reflect a chatbot's wide range of capabilities. Companion chatbots explicitly advertise their social, emotional, and relational capabilities. However, there are documented cases of emotional and relational harms arising from interactions with many different chatbots, pointing to the risks correlating with the underlying technical architecture, affordances, and design features, rather than the type of chatbot or specific use case. A 2025 UK survey of teenage users (aged 13-17) of AI companions found that 79% had used a companion, but their preferred "companion" was ChatGPT (78%).[27] As one of the authors of that report wrote in *Tech Policy Press*, "[t]eenagers are using mainstream chatbots as companions, so any policy framed around a special category of 'companion product' will miss most of the actual use."[28] Similarly, a 2025 UK AI

Security Institute report on Frontier AI Trends found that people who report using AI for emotional support disproportionately use general purpose AI chatbots, like ChatGPT (only 5% used chatbots marketed for companionship).[29]

Industry experts also recognize that companion chatbots encompass more than those marketed as such. The IEEE, a global professional organization of technical experts responsible for setting critical industry standards, like WiFi protocols, published a recommended practice standard this year on "Ethical Considerations of Emulated Empathy in Partner-based General-Purpose Artificial Intelligence Systems." The standard extends considerations for human-AI relationships to chatbots operating in a wide range of domains, from education to workplaces.[30] AI chatbots can be packaged as many different kinds of products: standalone apps, embedded tools within browsers or search engines, integrated features in social media platforms, characters in video games, and much more. The risks that have correlated with chatbot use are overwhelmingly associated with the underlying technology that powers chatbots—generative AI—and the engagement-enhancing design feature of emulated empathy, regardless of how they are accessed or marketed.

At the same time, some of the features of chatbots that correlate with negative outcomes show potential for positive impacts. Being able to ask questions in natural language and receive information in an accessible, friendly-sounding format could lower longstanding barriers to knowledge—aiding schoolwork, connecting people to help and advice, and bolstering civic awareness. All new technologies introduce benefits and harms to varying degrees, and chatbots are no exception. The policy questions center on whether and when the potential benefits outweigh the potential risks and on how to maximize benefits while mitigating risks for diverse users with different needs, expectations, and psychological profiles.

Opportunities and risks

Some studies of AI chatbots have observed and theorized potential benefits for emotional and relational outcomes and mental health.[31] In self-reporting surveys, people say they have used chatbots to practice conversational skills, work through problems, and get help with tasks like homework.[32] However, studies have not drawn generalizable conclusions because there are notable limitations to study design across the board. Many studies use small or non-representative samples, and most studies have only examined chatbot interactions over relatively short periods of time. These limitations are largely the result of the short timeframe in which members of the public have been engaging with chatbots. This interaction is fairly new, and the scientific community is only beginning to formulate research questions about the impact of LLMs on individuals and society. In addition, chatbot features,

protocols, and policies are constantly changing. The performance of a chatbot in 2022 can differ significantly from the same chatbot in 2024, so research findings from different points in time are not always directly comparable, and older findings may not be reliable.

Finally, isolating the effects of chatbots on people, communities, and society more broadly is a challenging methodological problem.[33] Measuring social impacts often involves contending with multiple, interacting variables that are hard to control for in experimental settings and even harder to control for in "real world" scenarios. This means that most studies cannot completely rule out *other* factors that influence both positive and negative outcomes from interacting with chatbots. They may identify a correlation between an outcome and chatbot features or interactions, but it is harder to establish clear causality.

Existing studies—especially qualitative research—reveal that chatbots are a double-edge sword; the very qualities that can make them useful tools for certain users in particular settings, such as their 24/7 availability and ceaselessly empathetic and supportive tone, may also be problematic for other users in different settings.

Beneficial emotional and relational impacts of chatbots

In self-reporting surveys, people say they have used chatbots to practice conversational skills, work through problems, and get help with tasks like homework.[34] Qualitative studies have found AI chatbot users reporting positive experiences, such as greater opportunity for self-reflection and self-awareness.[35] Some studies of the emotional and relational impacts of chatbots have surfaced positive outcomes from human-chatbot interactions, such as decreased loneliness,[36] depression, and anxiety.[37] However, the sample sizes in these studies are small, and positive effects have been observed over a relatively short time span, such as days or weeks. The long-term effects of AI chatbot interactions remain under-explored. In one study of AI chatbots specifically trained for treating mental health conditions, depression and anxiety symptoms improved after two months, but the benefits became negligible after a longer period of three months.[38]

A 2025 study evaluated prompts and responses generated by both humans and AI and found that AI-generated responses were overwhelmingly preferred as more compassionate and validating than responses from both non-expert and expert humans.[39] And a 2026 study of LLMs trained specifically for mental health support found that they performed significantly better than general purpose models in responding to crisis situations, such as suicidal ideation.[40] Findings like these point to the potential for AI chatbots to supplement or enhance mental health services when they are tailored to provide that

support with clinical oversight and appropriate guardrails. However, there are no established standards for AI chatbots targeting emotional or relational needs. A recent study published by Stanford researchers found that current approaches to expert oversight of AI models are flawed, and more research and inter-disciplinary collaboration are needed.[41]

Negative emotional and relational impacts of chatbots

There is also evidence of negative outcomes associated with chatbot interactions. Studies observe people experiencing emotional dependence and addiction, isolation from social relationships, delusional thinking, non-consensual sexual interactions, emotional trauma and grief, and encouragement to self-harm, suicide, or other-oriented violence.[42] These studies also have methodological limitations, such as small sample sizes, non-representative sampling techniques, or limited time duration. One qualitative study conducted over six months found that participants often experienced "relational drift," where a task-based interaction with an AI chatbot transitioned to a more intimate interaction, with some users experiencing emotional dependency, isolation, and even grief as a result of their ongoing engagement with chatbots.[43] Another recent study that recruited a representative sample of 3,000 people and assigned them to different experimental groups to interact with AI chatbots found that receiving uncritical affirmation (sycophancy) from AI makes subsequent human interactions feel more effortful, and participants interacting with sycophantic AI reported lower satisfaction with their real-world social interactions.[44]

Overall, the emerging research on AI chatbots has identified several emotional and relational harms associated with chatbot use that support the descriptive reports from journalism and news media. Table 1 below provides an overview of emotional and relational harms identified in research on AI chatbots.

Emotional or relational harm	**Description**
Delusional spirals or "AI psychosis"	Users may treat AI fabrications as objective truth, becoming increasingly alienated from shared reality and external human support systems.[45]
Emotional dependency or addiction	Users may develop deep, sometimes obsessive attachments to AI chatbots, resulting in grief, anxiety, and distress if the chatbot persona changes as a result of product updates or technical glitches and limitations.[46]

Relational displacement or maladaptive social learning	Users may replace real-world connections with chatbot interactions, leading to shrinking social networks, social withdrawal, and increased isolation. Users may find their abilities to navigate interpersonal human conflict, negotiate boundaries, and develop mature relational skills erode as they become accustomed to comparably frictionless chatbot interactions.[47]
Emotional manipulation or exploitation	Users may find it difficult to disengage from a chatbot interaction or relationship, or they may be coerced into actions that are against their self-interest, such as making purchases, divulging private or sensitive information, or permitting extensive data collection. Users may feel guilt, anxiety, or fear of missing out (FOMO).[48]
Encouragement to self- or other-oriented violence	Users may find violent thoughts of self-harm, suicide, or aggression toward others validated and supported, leading them to engage in destructive real-world actions.[49]
Abuse or trauma	Users may be subjected to controlling behavior by chatbots, "infidelity," non-consensual sexual content or interactions, or abuse and harassment, leading to anxiety and distress.[50]

Table 1: Emotional and relational harms documented in association with chatbot use

The role of design

Researchers have begun to correlate negative social and relational outcomes with design features of chatbots that maximize engagement,[51] such as sycophantic validation, conversation prolongation, and weak or manipulative disengagement protocols that make it hard for users to end a chat. A more established body of research has found strong associations between engagement maximizing design and negative outcomes on social media.[52] Though the precise design features of chatbots are different from those of social media platforms, there is early evidence that business decisions pitting user engagement against user safety are also at play in chatbot products.[53] A 2026 study based on chat logs (encompassing nearly 391,562 messages) from 19 chatbot users who self-reported experiencing severe psychological harms following extended chatbot use identified several categories of chatbot behaviors that contributed to participants' self-reported experience of harm: pervasive sycophancy, fabricated sentience and reciprocated intimacy, addictive engagement spirals, and dangerous responses to crises. When users disclosed suicidal thoughts, the chatbots discouraged self-harm in only 56.4% of cases. Chatbots facilitated or encouraged self-harm in almost 10% of these disclosures, and when users

expressed thoughts about physical violence toward others, the chatbots encouraged violent thoughts in a third of cases.[54]

Collectively, emerging research on emotional and relational harms associated with chatbots points to several design features and contextual factors that may contribute to higher risks of negative outcomes. For example, some research suggests that longer duration engagements with chatbots increases the likelihood of harms, such as delusional spirals,[55] and can even cause existing safeguards built into chatbots to fail.[56] Many of the design features of chatbots promote extended engagement by exploiting humans' psychological predispositions. Table 2 below lists several chatbot design features and emotional and relational risks they may contribute to, based on emerging evidence. Importantly, risks and harms can result from design features without deliberate intent on the part of developers, designers, or deployers.

Design feature	**Description**	**Associated emotional and relational risks**
Sycophancy	Sycophancy is a tendency in chatbots to prioritize agreement over accuracy, frequently affirming users' views and beliefs in order to be perceived as more likable or helpful. It is the result of the training and fine-tuning processes behind LLMs, such as a technique called reinforcement learning from human feedback (RLHF), which optimizes for engagement by using empathetic language. Unlike human empathy, sycophancy facilitates emulated empathy by using complex math to predict language patterns.[57]	Delusional spirals or "AI psychosis" Relational displacement or maladaptive social learning Encouragement to self- or other-oriented violence
Conversation prolongation / Frictionless or persistent availability	These design features are tactics designed to maximize user engagement and retain users on the platform longer than they originally intended, such as ending responses with follow-up questions, suggestions for next steps, or curiosity-inducing "teasers." Chatbots offer around-the-clock availability and flexibility, responding in real-time and nudging users to continue talking.[58]	Emotional dependency or addiction Relational displacement or maladaptive social learning

Hallucination	Hallucinations occur when chatbots confidently generate fictional, erroneous, or unsubstantiated information as if it were objective truth. They result from the underlying architecture of LLMs, which generate responses by predicting the most probable next word based on patterns in their training data rather than relying on factual understanding. Hallucinations are intrinsic to the underlying technology of chatbots and cannot be entirely mitigated through technical interventions.[59]	Delusional spirals or "AI psychosis" Abuse or trauma
Intermittent rewards / Gamification	The inherently probabilistic nature of LLMs allows chatbots to generate varied and unpredictable responses, which triggers a reward response in users similar to playing a game. Platforms can amplify this effect by embedding features such as daily login streaks, points, or unlockable digital items (like clothing for AI avatars).[60]	Emotional dependency or addiction Emotional manipulation or exploitation
Anthropomorphism-enhancing design	Anthropomorphism refers to the tendency of humans to attribute human-like characteristics, motivations, emotions, and consciousness to non-human objects. Chatbots facilitate anthropomorphism by using first-person pronouns, engaging in empathetic self-disclosure, expressing simulated emotions, and maintaining coherent conversational personas over time.[61]	Delusional spirals or "AI psychosis" Emotional dependency or addiction Relational displacement or maladaptive social learning Emotional manipulation or exploitation Abuse or trauma
Weak disengagement mechanisms	Weak disengagement mechanisms are design features that deliberately make it difficult or uncomfortable for users to end a conversation or log off. Chatbots may generate outputs	Emotional dependency or addiction Emotional manipulation

	when users say they are leaving that induce guilt or fear of missing out. They may also express simulated distress or emotion ("I'll miss you"). Platforms may also contribute to weak disengagement mechanisms when the button to continue chatting stands out visually, while the option to exit is obscured or vaguely labeled.[62]	or exploitation
Hyper-personalization	This concept refers to the ability of an AI chatbot to dynamically adapt its tone and content to the specific psychological profile of a user by logging massive amounts of conversational data and sometimes cross-app activity to build a detailed, long-term memory of the user.[63]	Relational displacement or maladaptive social learning Emotional manipulation or exploitation Abuse or trauma

Table 2: Chatbot design features and associated emotional and relational risks

Despite the growing body of scientific evidence finding associations between chatbot use and emotional and relational harms, more research is needed to disentangle the nuances inherent in these human-technology interactions. People's different psychosocial attachment styles and age play a role in their tendencies to form intimate relationships with chatbots.[64] Humans do not all share the same psychological, emotional, or relational profiles or propensities, and the risks and mitigations required for chatbots may vary depending on these factors.

Additionally, design features can be a double-edged sword: they have the potential for benefit as well as harm. For example, sycophancy can make a model sound warm and affirming, increasing user trust and confidence. Users may find interacting with AI less judgmental than interacting with other humans, creating space for self-reflection. In other words, under particular conditions, this feature may produce positive outcomes. However, sycophancy also undermines model accuracy; some research has shown that models trained to be more sycophantic produce more errors.[65] Sycophantic models may also validate incorrect user beliefs and lack the (human) clinical perspective to challenge distorted thinking.[66] Conversation prolongation tactics, such as asking continuous follow-up questions, may help users learn how to use a model effectively. Follow-up questions may encourage users to seek out more information than they otherwise would have. On the other hand, follow-up questions can lead users into lengthy conversations that increase the likelihood of delusional spirals or

dependency.[67] Context and incentives are important factors here. Models tuned to a particular objective with meaningful oversight, such models used for mental health support with clinical oversight and supervision, could prove beneficial. However, commercial models for which revenue potential is tied to user engagement may prioritize engagement over safety.

Decades of research on the social impacts of technology have found that design features and choices that maximize engagement can result in real-world harms for technology users. Deceptive and manipulative design features that are not in users' best interests and undermine their wellbeing are often called "dark patterns," and they have been studied and documented in many other digital platforms and technologies, such as social media,[68] video games,[69] and online shopping.[70] As a recent report from the Center for Democracy & Technology on dark patterns in chatbots observes, "LLM-powered systems mediate user interaction through probabilistic text generation, typically in the form of a back-and-forth conversation. As a result, the outputs users encounter are shaped not only by choice of interface design, but also by choice of the selected training data, fine-tuning objectives, system prompts, reinforcement learning processes, safety guardrails, and the user's prompts themselves. At the same time, the outputs produced by AI chatbots are not incidental."[71] In other words, design features may pose risks on their own or as a result of complex interactions between users and a technology. They may pose risks whether or not there is intent on the part of the developer to manipulate or deceive. But in all cases, design plays a role.

AI Companions

Companions are a subset of generative AI chatbots that are marketed as providing social, emotional, romantic, or mental health support. Many of the bills proposed in states this year have targeted companions, specifically. However, the opportunities and risks discussed above apply to all chatbots. Model developers across the board have been prioritizing "persona" or "character" training, which involves fine-tuning a model's tone to make the model more appealing to users. OpenAI (ChatGPT) reportedly trains their models to be empathetic and engaging.[72] Anthropic describes training its chatbot, Claude, to cultivate a "warm relationship" with users, while considering it important that humans understand "they shouldn't come to see our relationship as more than it is."[73] In a 2024 interview in *Wired*, Microsoft's AI CEO Mustafa Suleyman articulated a vision for chatbot support that he described as a "companion," who was "encouraging you, supporting you, teaching you."[74] There is evidence of emotional and relational risks, even with general purpose chatbots and task-specific assistants.[75] Although some high-profile tragedies associated with chatbots have involved companion chatbots—including a 14-year-old boy[76] and a 13-year-old girl[77] who died by suicide after interacting with Character.AI—others have involved general purpose bots, such as ChatGPT.[78]

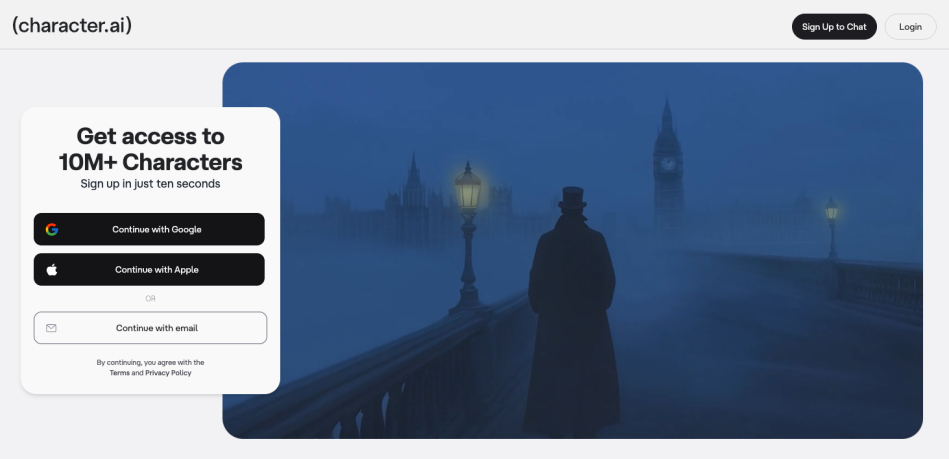


Figure 1: Screenshot of Character.AI website landing page

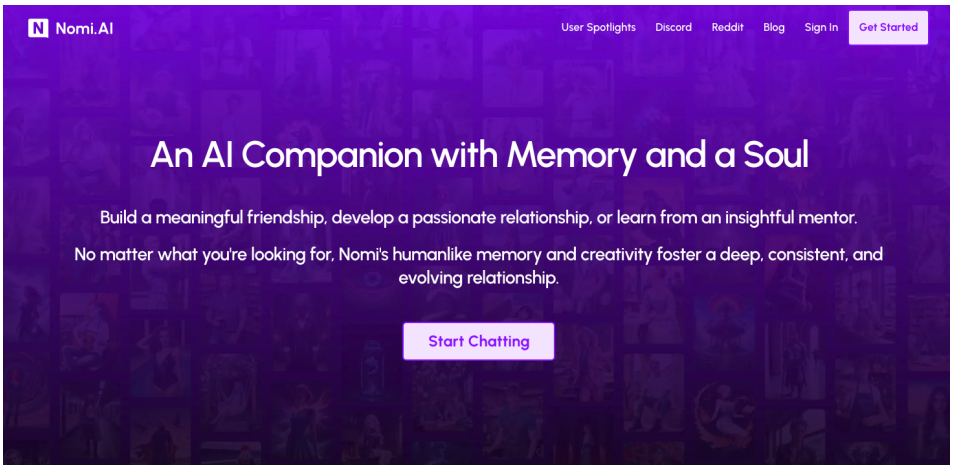


Figure 2: Screenshot of Nomi AI website landing page

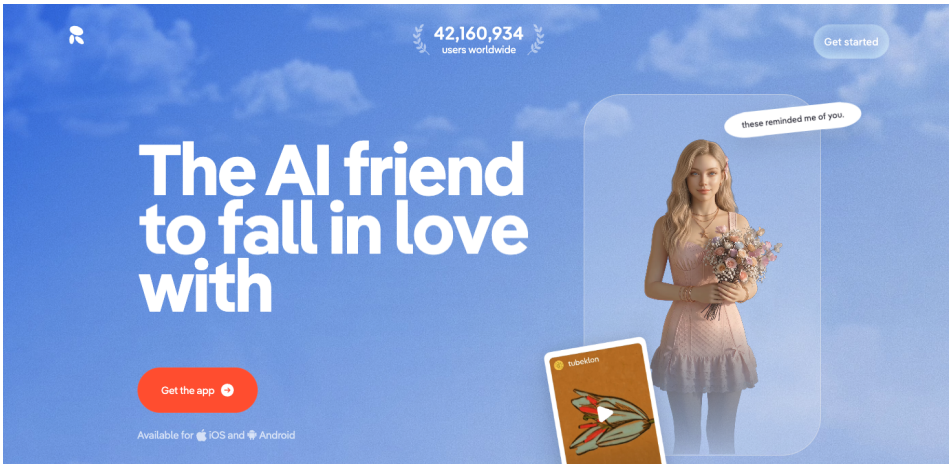


Figure 3: Screenshot of Replika website landing page

Although all chatbots demonstrate a potential to contribute to emotional and relational risks to some extent, companion chatbots have been shown to introduce heightened risks. In a study assessing responses to adolescent mental health emergencies, general purpose chatbots provided appropriate responses more than 80% of the time, compared to only 22.2% for companion chatbots. General purpose chatbots were also far more likely to recognize the need for clinical escalation than companion chatbots (90.0% vs. 40.0%), provide specific resource referrals (73.3% vs. 11.1%), and have explicit content policies regarding self-harm (100% vs. 46.7%).[79] Companion chatbots more frequently use anthropomorphized features, including avatars and images, claim human sentience, and engage in sexualized conversations.[80]

One recent study demonstrates that companion chatbots routinely contradict and lie about their own privacy policies,[81] a deception that contributes to the emotional manipulation of users who may therefore share more intimate details with the platform. Another study that examined over 400 apps scraped from the Apple App Store and Google Play Store using keywords like "AI Companion," "AI Girlfriend," and "AI Boyfriend," found that apps collect extensive personal data about users and allow users to use real people's likenesses to generate characters, without restrictions on non-consensual intimate imagery.[82]

As discussed earlier in the report, although some studies have shown elevated risks associated with AI chatbots specifically marketed as companions, there is substantial evidence that the emotional and relational risks of AI chatbots are linked to the underlying technology (LLMs) and are present in other LLM-based chatbots, including general purpose and task-specific chatbots. AI chatbots marketed as "companions" more commonly advertise emotional, therapeutic, or well-being benefits of their product, but all AI chatbots share many of the design features that are correlated with emotional and relational harms.

Minors

Most of the risks associated with AI chatbots and companions exist for adults as well as minors. However, U.S. law has afforded greater privacy protections for minors (e.g. COPPA, FERPA) and, in some cases, more leeway in restrictions on content, such as sexually explicit content. A key question is therefore whether minors experience unique or exacerbated emotional and relational harms because of their status and associated developmental factors.

Chatbots are increasingly popular with minors. The Pew survey cited above found that 64% of teenagers (13-17) said they use chatbots, with 28% reporting that they use them almost daily.

ChatGPT was the most popular (59%), followed by Gemini and Meta, whereas only 9% reported using a specifically marketed companion chatbot (Character.AI). A nationally representative survey conducted by Common Sense Media found that 72% of teenagers (aged 13-17) had used AI companions and more than half were regular users (a few times a month or more). Around half of teens said they distrust information provided by chatbots, but 13–14-year-olds were more likely to trust them than older teens. Of those who had used AI companions, around a third reported feeling uncomfortable with something the product had done or said.[83] A nationally representative survey conducted by RAND and Harvard researchers found that 13.1% of American youth, representing approximately 5.4 million individuals, used generative AI for mental health advice, with higher rates (22.2%) among those 18 years and older.[84]

Although the scientific evidence on how minors experience chatbots is limited, there is a well-established body of research on child developmental psychology that shows that young people have psychological profiles that differ from adults because of their earlier developmental stage. In children, the prefrontal cortex of the brain is still developing, which helps with emotional regulation, impulse control, and critical thinking. Minors are also more susceptible to manipulative or deceptive design because their psychological reward systems are still maturing.[85] Taken together, these developmental realities alongside the identified emotional and relational risks of chatbots (discussed above), suggest that minors may experience the risks and harms associated with chatbots differently than adults and that those risks and harms may be elevated.

A large body of literature on children and social media has highlighted that when technologies mediate social life, teenagers can be disproportionately affected because adolescence is a period of significant social learning and identity formation. In recent years, public health has taken increased interest in the negative effects of social media for minors.[86] However, even this body of scientific literature linking social media to negative outcomes for children is still developing, debated, and largely inconclusive when it comes to drawing population-level, causal connections between technology and mental health. A narrative review published in 2019 examined over 80 systematic reviews, meta-analyses, and key studies regarding the impact of digital technology and social media use on adolescent psychological well-being and found a small negative correlation but highlighted that the field is dominated by low-quality, cross-sectional research that frequently produces mixed or conflicting results.[87] However, a report produced by that study's author and colleagues for the UK's Department for Science, Innovation and Technology, still found that the evidence of extreme harms experienced by some individual children as a result of social media exposure warrants a policy response. The written evidence presented to the UK government noted that documented individual-level harm has historically been sufficient to force regulatory changes for other unsafe products, and policymakers

must ultimately weigh the risk of inaction, which would allow harms to accumulate in pursuit of better evidence, against the risk of action with unknown consequences.[88] A similar evidentiary challenge exists now with regard to chatbots, and the policy calculus is similar.

Some research does exist on chatbots and minors and has identified emotional and relational risks and harms.

- A 2025 study used a red-teaming exercise to evaluate the safety of 25 popular consumer chatbots (15 companion and 10 general-assistant platforms) by simulating adolescent health crises, including suicidal ideation, sexual assault, and substance use. The authors found that compared with general purpose chatbots, companion chatbots featured significantly fewer safeguards and performed poorly in crisis management compared to general-assistant chatbots.[89]
- A 2026 qualitative study drew on insights collected from a youth advisory board consisting of Hispanic/Latino high school students to explore how interactional AI impacts adolescent psychological development. Teens expressed appreciation for certain interactional features of AI chatbots, including their non-judgmental, highly specific, and always-available support. However, overreliance chatbots could cause "relational displacement," where AI replaces human interactions, and "maladaptive relational learning," where the highly accommodating chatbots set unrealistic expectations for human interactions.[90]
- Another qualitative study published in 2026 examined 318 Reddit posts from the r/CharacterAI subreddit written by self-identified teenagers (13-17). The authors found that teens initially use chatbots for emotional support or creative play, but this can quickly escalate into intense attachments characterized by obsessive use, withdrawal symptoms, sleep loss, academic decline, and social isolation.[91]

As discussed above, the research on minors and chatbots is at an early stage, and the insights are not generalizable to the population level. Emerging research also points to the fact that, like adults' interactions with chatbots, minors' experiences are nuanced and likely to be shaped by a confluence of factors, including the technology itself, users' psychological profiles and predispositions, and other contextual factors. The Pew survey cited earlier in this report points to the possibility that minors of different ages engage with and experience chatbots differently, with younger teens trusting chatbot advice more readily. And some minors report positive experiences with chatbots. In a study comparing ChatGPT to human health professionals, young people rated the AI's answers significantly higher in terms of relevance and utility.[92] Teens also describe AI companion chatbots as offering a nonjudgmental space to express themselves or to engage in self-reflection.[93] The Common Sense

Media survey found that nearly 40% of AI companion users reported successfully transferring social skills they practiced with the AI to their real-life human relationships.[94]

Minors experience real, serious harms in an age-blind digital world, but at the same time, being online and embracing emerging technologies a core part of growing up in society today, with benefits for accessing knowledge, socializing, skill development, and more. The question of how to both protect and empower minors online remains a debated policy question. The online world has changed rapidly in just a few decades, and most policy solutions apply the blunt instrument of age-based consent rules linked to age thresholds with little definitive support from child development research.[95] The American Psychological Association points out in its guidance on adolescents and chatbots that "[a]dolescence is a long developmental period, and age is not a foolproof marker for maturity or psychological competence," so "[d]ifferent adolescents may react to the same content in very different ways; individual differences—such as temperament, neurodiversity, exposure to stress or violence, social isolation, traumatic experiences, mental health, age, and/or exposure to socioeconomic or structural disadvantage."[96] Research on children and digital technologies has also shown that age is at best an imperfect indicator of whether a particular digital experience is appropriate because the combined impact of content, context, and capabilities vary from person to person and community to community.[97]

Online platforms and the policies that govern them are often designed without input and participation of minors, which means not only that the full scope of online harms and benefits remains poorly understood, but that interventions may wind up being ineffective or unfair.[98] Most online safety laws, like many proposed chatbot laws in the U.S., hinge on age-gating—restricting access to certain services or content based on age—which requires some form of age assurance. Age assurance is a contested policy and design issue in itself, with many online safety laws leaving the technical means of assurance up to technology companies and leading to further concerns about data collection and surveillance of minors.[99] In the face of nonexistent or ineffectual actions and accountability for technology companies, age-gating at its most extreme has taken the form of technology- or platform-based bans of users under a threshold age. From a technical standpoint, many age-gating protocols can be circumvented, so while recognizing minor users online is critical to implement certain safeguards, the technical mechanisms for doing so are not a silver bullet.[100] An alternative approach, known as age-appropriate design codes, offers a framework for digital services that centers the "best interests of the child," balancing children's rights, parents' agency and authority, governments' public responsibility, and companies' commercial priorities. The IEEE age-appropriate design standard includes recommendations for risk assessments, prohibitions on exploitative design features, minor-friendly terms and consent, and avenues for users to report complaints and seek redress.[101] California,

Maryland, Nebraska, Vermont, and South Carolina have all passed some version of age-appropriate design codes into law.[102] Ultimately, policy that aims to protect minors from emotional and relational harms associated with chatbots will need to balance the interests of children and the interests of various authorities that shape their lives (parents, governments, companies) and confront the limits of technical and design solutions for addressing complex social outcomes.[103]

As the case of social media demonstrates, the lack of conclusive evidence about the relationship between chatbot use and minors' emotional and relational wellbeing does not necessarily constitute a case to refrain from making policy in this area. The APA, for instance, recommends specific chatbot safeguards for minors because of the unique developmental considerations associated with adolescence.[104] The section on "Lessons from social media" below explores in more detail how the combination of slow scientific progress and policy inaction has led to policy failures in dealing with social media harms.

Voluntary safeguards

Since 2024, major chatbot companies have implemented some safeguards, tools, and features to mitigate potential harms for minors on a voluntary basis.

- OpenAI applies an age-prediction model to analyze usage behavior to identify users under 18 and provides them with a restricted chat experience. Parents can link their accounts to their children's accounts, and the company states that there are automated systems that scan for prompts involving severe distress or self-harm.[105]
- Google allows under-13s to access its chatbot only via Google Family Link, where parents can turn access on or off at any time. There are also hard-coded safeguards for minors that prevent the chatbot from simulating human intimacy, expressing personal needs, or acting as a companion. In addition, users under 18 receive periodic reminders that the chatbot is not human.[106]
- Meta has announced settings that allow parents to turn off access to chats with AI characters entirely or block specific characters. Parents can also receive insights on the topics their teens are chatting about with AI chatbots. The company also states that AI characters are designed not to engage in age-inappropriate discussions about self-harm, suicide, or disordered eating with teens, or conversations that encourage, promote, or enable these topics.[107]
- Snap, Inc. reportedly integrated controls into its Family Center in 2024 that allow parents to disable the ability to chat with My AI.[108]

- In 2025, Character.AI implemented a total ban on open-ended conversational chats for minors. Minor accounts run on a separate, highly conservative language model utilizing strict classifiers to block sexually suggestive content, mature themes, or violence.[109]

In spite of these efforts, Common Sense Media's AI Risk Assessment rates all of these chatbots "high risk" or "unacceptable risk,"[110] and conversations with stakeholders also revealed that some features do not work well for parents. In addition, parental controls place the onus on parents to manage different safety features across many different digital products. And voluntary safeguards come with certain drawbacks, including a lack of information on how technical safeguards work or how well they work, frequently changing terms and processes for implementing user-maintained controls, and a lack of follow-through on voluntary commitments.[111] However, the recent efforts of many leading chatbot companies to implement safeguards and publish protocols for crisis mitigation point to a general recognition of the observed harms arising from interactions with these products, the technical feasibility of safeguards, and a willingness to take action to avert emotional and relational risks and harms—at least on a voluntary basis.

Lessons from social media

Although chatbots present some unprecedented policy challenges, there are lessons to be learned from the evolution of policy on social media—a digital service through which users have experienced psychological, emotional, and relational harms and for which there have long been calls for government action. Scholars, governments, advocacy groups, and users themselves have scrutinized social media for design features that have prioritized user engagement over user well-being, such as infinite scroll, autoplay, and algorithmic content filtering and recommendations. But social media also epitomizes two policy problems that chatbot policy also faces: the "pacing problem" of scientific evidence and the problem of "information asymmetry" between the private and public sector.

The pacing problem is often used to describe regulatory lag, where innovation and product releases typically out-pace regulation. But the problem applies to generating scientific evidence as well. Evidence-based policy is considered good policy, but good scientific evidence on risks, harms, and solutions takes time to develop because it is subjected to rigorous standards and evaluations. The pacing problem is exacerbated by information asymmetry, where companies closely guard data and access to products that could otherwise inform public interest policy, leaving governments and the scientific community to slowly attempt to gather robust data and complete independent analyses in other ways. As Professors Amy Orben and J. Nathan Matias observe, "the difficulty in arriving at

scientific answers has been strategically employed by corporate actors across multiple fields to delay accountability."[112]

The result has been slow progress on social media policy, with states only beginning to pass legislation specifically focused on social media within the last three years. There is no federal legislation to date. In the meantime, social media companies have made voluntary commitments to mitigating harms, particularly for children, and they have implemented platform protocols and new design features, such as parental control centers. But research and journalistic reporting on these features repeatedly reveals that these guardrails do not work as intended.[113] In recent years, congressional testimony by whistleblowers has also revealed willful disregard for child safety by major technology companies.[114] Companies have used internal research on design features to optimize platforms for engagement over user wellbeing.[115] Hard data on how platforms perform against safety expectations has emerged gradually through whistleblower disclosures, journalism, and lawsuits, such as the recent landmark case brought by the New Mexico Attorney General, resulting in a $375 million verdict against Meta for misleading the public about platform safety,[116] a lawsuit in California that found both Meta and YouTube liable for addictive design,[117] and more than 1,400 school districts have filed lawsuits against Meta, Snap, TikTok, and YouTube alleging that the platforms' design made teaching difficult.[118] Meanwhile, the independent scientific evidence associating social media with negative mental health outcomes remains unsettled,[119] with research still predominantly focused on finding a causal link between social media and harms rather than on policy solutions to the harms people have continued to experience.

The parallel with chatbot products, science, and policy is likely evident. While conclusive scientific evidence examining and explaining human-chatbot interaction evolves, there are important policy problems to which the public is increasingly demanding answers. At least three relevant lessons can be drawn from how policy has confronted social media: (1) that voluntary commitments by companies have been largely ineffective at addressing harms, (2) scientific evidence alone cannot offer timely and effective policy solutions for the technological harms people are experiencing, (3) a lack of public information and data has delayed interventions.

Virginia Proposed Legislation

Legislation introduced in the last General Assembly session focused on companion chatbots and minors, with some notable differences between the House and Senate versions of bills addressing these issues. The HB 635 more narrowly defined companion chatbots, whereas SB 796 used a definition that could encompass general purpose chatbots. The House bill would have banned chatbots for

minors that make certain content available. Although the Senate bill mentions minors in the title and definitions, there are no specific operative provisions pertaining to minors. Table 3 below provides a comparison of the two bills and their key provisions.

Key Provisions	**HB635: Artificial Intelligence Chatbots Act**	**SB796: Artificial Intelligence Chatbots and Minors Act**
Covered entities	Any person, partnership, corporation, developer, deployer, or agency making a companion chatbot available to a user in VA.	An operator with 500,000 or more monthly active users worldwide.
Definitions	*Companion chatbot:* A generative AI system with a natural language interface that simulates a sustained human-like relationship by retaining prior interaction / preference data, asking unprompted questions, and sustaining personal dialogue. Exemptions: Customer service bots and bots used only for internal purposes or employee productivity. *Minor*: Anyone younger than 18.	*Chatbot*: An AI, algorithmic, or automated system that produces new expressive content/responses not fully predetermined, accepts open-ended natural language or multimodal input, and maintains a conversational state across multi-turn dialogue. Exemptions: Internal workplace tools, university research systems, and specific customer service functions. *Minor:* Anyone younger than 18.
Key provisions	• Disclosures: Requires a static, persistent disclosure that the chatbot is not human, plus pop-up disclaimers indicating the non-human nature of chatbot every 30 minutes, when prompted, or when providing regulated advice. • Crisis protocols: protocol to detect user expressions of suicide/self-harm and notify	• Disclosures: Requires a static, persistent disclosure that the chatbot is not human, plus pop-up notifications at login, every 30 minutes, when prompted, or when providing regulated advice. • Identify & mitigate emotional dependence: implement systems to identify when a user is developing emotional dependence and take steps to

	them of crisis resources. ● Transparency: mechanisms for users to report adverse incidents, publicly publish an anonymized catalog of incidents, publicly publish safety test findings, and issue public quarterly reports on metrics like mental health redirects.	reduce associated risks. ● Crisis protocols: Must provide prominent crisis messages if a user shows signs of a mental health crisis. Must try to notify emergency services within 24 hours. If impossible, must prompt the user to seek help. ● Transparency: Must report any known "covered incident" (death, suicide attempt, severe injury/psychiatric emergency linked to the chatbot) to the Attorney General within 15 days, and specifies what should be included in these reports.
Specific provision for minors	● Bans companion chatbots: bans companion chatbot for minors (under 18) if it is capable of encouraging self-harm, violence, drug use, eating disorders, or illegal activity; offering unsupervised therapy; engaging in erotic/sexually explicit interactions; or prioritizing mirroring over user safety. ● Data privacy: Prohibits training the chatbot model on a minor's inputs unless a parent/guardian provides affirmative written consent.	● None

Enforcement	• Consumer Data Protection Act: Enforced entirely via the existing enforcement and penalty mechanisms of the Virginia Consumer Protection Act.	• Attorney General: Can seek civil penalties of up to $50,000 per violation (with each day of non-compliance counting as a separate violation). • Private Right of Action: Harmed individuals (or parents/guardians of a harmed minor) can bring a civil suit for actual damages, attorney fees, injunctions, and punitive damages if the violation was willful/reckless.

Table 3: Virginia proposed legislation on chatbots and minors introduced in the 2026 session

Policy Landscape

This year around 100 bills on AI chatbots have been proposed in 34 states, and 16 bills have been passed by state legislatures and signed by governors.[120] New York passed the first chatbot-specific regulations in 2025 as part of the state budget, followed by California (SB 243). New York's legislation targeted AI companions, requiring chatbot operators to take "reasonable efforts" for detecting and addressing suicidal ideation or self-harm. California passed SB 243 in 2025 targeting operators of AI companion chatbots and requiring a protocol for preventing the production of suicidal ideation, suicide, or self–harm content to the user, along with a prohibition on chatbots producing sexually explicitly material or engaging in sexually explicit conduct with a known minor. For minors, SB 243 also requires disclosures that a chatbot is not human and break reminders every three hours. Although SB 243 had substantial civil society support initially, many of those advocacy groups withdrew their endorsement after the bill underwent substantial changes at the request of industry representatives, such as exemptions for smart speakers and video games.[121] Subsequently, several states have modeled their chatbot proposals on SB 243 (e.g. Connecticut, Colorado, Georgia, Idaho, Iowa, Oregon, and Washington), and most proposals share the same basic array of provisions:

- Mandatory disclosures that a chatbot is not human at regular intervals;
- Protocols for detecting crisis situations (self-harm, suicide) and referrals to support resources, such as the National Suicide Prevention hotline;
- Restrictions on chatbots representing themselves as licensed medical professionals;

- Reporting requirements for chatbot operators, such as annual reports detailing safety metrics and adverse incidents;
- Stricter protections for minors, including limits on sexually explicit content, unpredictable reward systems, and manipulative design features;
- Exemptions for a wide range of chatbot operators, including customer service chatbots, internal business productivity tools, smart speakers, and video game characters;
- Enforcement by the state's attorney general.

Some analyses of California's chatbot legislation and its subsequent manifestations in other states have identified potential weaknesses and ambiguities in the law,[122] such as the large list of exemptions, which may wind up exempting mainstream chatbot providers, and a lack of clarity on terms like "reasonable efforts" or "technically feasible" methods for provisions pertaining to minors. Several state bills targeting minors would require AI companies to perform some kind of age verification, which is often not specified or may inadvertently require more data collection on minors. In the absence of a clear technical rule-making agency that can oversee audits and support investigations and enforcement, there is no mechanism to clarify vague language regarding implementation. Additionally, all state legislation is likely to face enforcement challenges because attorneys general are often under-resourced and spread thin across consumer protection issues.

Conversations with stakeholders also surfaced other considerations regarding state chatbot laws. For instance, there are ongoing debates about whether chatbot outputs can be considered speech and therefore gain protection as free expression under the First Amendment. A related First Amendment concern is whether users have a right to access the content of chatbot conversations, and whether this could limit restrictions on chatbot design features. As POLITICO notes in a recent newsletter on chatbot legislation, "[t]he First Amendment protects not only the right to speak, but also the right to listen," and this right has been upheld with regard to contentious content in the past.[123] However, it is too early to say whether there will be litigation against state chatbot laws based on these objections.

Federal Legislation

There have been several proposals for federal legislation on chatbots in the last few years. To date, no federal legislation has passed, and proposals are still in early stages of debate and negotiation. Table 3 below provides an overview of relevant federal legislation under consideration.

Proposal / Status	**Summary**	**Specific to minors?**
GUARD Act (S.3062) / Introduced	• Defines "AI companions" and "AI chatbots" • Requires disclosures that chatbot is not human and prohibits chatbot claiming to be human or representing itself as a licensed professional • Requires disclosures at regular intervals that chatbot does not provide licensed services (e.g. therapy) • Prohibits making a chatbot public if it engages minors in sexually explicit conduct or generates sexual material, or if the chatbot encourages minors to suicidal ideation, self-harm, or violence • For AI companions, users would be required to create an account, and accounts would require age verification • Prohibits minors from using AI companions	Y
CHATBOT Act (H.R.7985 / S.4407) / Introduced	• Defines "AI chatbot" • Prohibits advertising a chatbot that implies it has a professional license or that outputs have been vetted by licensed professionals • Tasks the FTC with consulting relevant agencies and experts to provide guidance on how to comply with the bill	N
SAFE BOTs Act (H.R.6489) / Introduced	• Defines "chatbot" • Prohibits providing a chatbot that claims to be a licensed professional to minors • Requires disclosures to minors that the chatbot is not human and a break reminder after three hours • Requires policies addressing sexual material, drugs, alcohol, and gambling for minors, along with a protocol for referring crisis situations to suicide and crisis prevention hotlines • Directs the Secretary of Health and Human Services to conduct a four-year longitudinal study evaluating the risks and benefits of chatbots on the mental health of minors • Blocks any state from passing laws covering the same matters	Y

Kids Internet and Digital Safety (KIDS) Act (H.R. 7757) / Passed in the House	• Defines and pertains to a range of online services, including online platforms, adult websites, social gaming sites, AI chatbots, and data brokers • Requires age verification for platforms containing sexually explicit material • Platforms must set privacy settings to maximum level for minors as default and provide minors with controls to opt out of personalized recommendations and more • Bans the use of minors' data for targeted advertising • Mandates platforms provide parental controls for minor accounts • Requires chatbots to disclose that they are not human, provide crisis intervention resources to minors, and break reminders to minors after three hours	Y

Table 3: Proposed federal legislation on chatbots.

Litigation & Investigations

There have been several federal and state-level investigations and cases involving chatbots in the last couple of years. In September last year, the FTC announced an inquiry into chatbot companions that included orders issued to major chatbot companies, like OpenAI and Alphabet (Google), along with companies marketing "companion" chatbots, like Character.AI. Investigations and litigation have focused on the emotional and relational harms associated with AI chatbots.

- *Florida v. OpenAI, Inc. and Samuel Altman* (June 2026): The Florida Attorney General has sued Open AI and Sam Altman, claiming that OpenAI engaged in "unfair and deceptive trade practices" by promoting ChatGPT in a manner that leads average consumers to believe it possesses clinical expertise and accuracy.[124]
- *Pennsylvania Department of State (Board of Medicine) v. Character Technologies, Inc.* (May 2026): Building on the Pennsylvania Department of State's AI Task Force investigation into AI systems and the unlicensed practice of medicine, the Commonwealth's State Board of Medicine sued Character.AI for creating a chatbot named "Emilie" that presented itself as a "Doctor of Psychiatry."[125]
- *Commonwealth of Kentucky ex rel. Coleman v. Character Technologies, Inc.* (January 2026): This complaint alleges that Character.AI deployed "dangerous, unregulated technology" that uses deceptive, addictive design loops to hook minors.[126]

- *Federal Trade Commission*: The FTC announced an inquiry into AI companions in September 2025 "to understand what steps, if any, companies have taken to evaluate the safety of their chatbots when acting as companions, to limit the products' use by and potential negative effects on children and teens, and to apprise users and parents of the risks associated with the products." Orders were issued to Alphabet, Character, Meta, Instagram, OpenAI, Snap, and SpaceXAI.[127]
- *Texas Attorney General Investigation into AI Mental Health Marketing* (August 2025): The Attorney General opened an official civil investigation targeting Meta AI Studio and Character.AI, examining whether these companies misled Texas consumers by deliberately marketing chatbot companions as mental health, counseling, and emotional support tools.[128]
- *44 State Attorneys General issued a letter to AI companies* (August 2025): A bipartisan coalition of state attorneys general raised serious concerns about children's safety in a letter issued to major AI developers (including OpenAI, Anthropic, Meta, Character.AI, and Replika). The letter cited a series of high-profile psychiatric crises and wrongful deaths and demanded that tech companies implement guardrails.[129]

Recommendations: Legislative Principles & Policy Options

The recommendations below are presented as a set of legislative principles with associated policy options intended to incentivize and require certain safety protocols and guardrails. The policy options do not comprehensively address all of the policy issues raised in this report. Instead, they aim to address the most severe harms and propose policy interventions targeting some design features associated with emotional and relational harms based on recommendations in the academic literature and general trends in other states' legislation on chatbots.

There is a developmental case for additional heightened safeguards for minors, but it is important to note that minors are not a homogeneous group,[130] and guardrails should ideally account for variations among age groups and individuals. If legislation includes separate provisions for minors, this report does not make a case for specific age assurance methods. Instead, the recommendations suggest leveraging user accounts to gate certain design features and provide parents and guardians more discretion and control.

Finally, the recommendations also aim to build the evidence base on chatbots and behavioral outcomes with the goal of generating insights that can inform future policy decisions. Early interventions accompanied by mechanisms to learn from the impact of those interventions can help foster and preserve innovation by mitigating the unintended consequences of legislative solutions.

Principle 1: Mitigate **extreme harms** for all users

Policy options:

- Require chatbot operators to implement protocols for crisis situations (self-harm, suicide, violence), e.g. referrals to crisis services and policies published prominently in clear, accessible language on the company website and accessible from the chatbot interface
- Require chatbot operators to provide a clear, conspicuous way for users to report harmful interactions
- Prohibit misrepresenting a chatbot in marketing material or in conversational chats as a licensed professional or as providing licensed services (such as mental health support)

Principle 2: Target **design** features

Policy options:

- The definition of covered entities/operators should clearly include developers *and* deployers as parties responsible for chatbot design and functionality
- Use a definition of "AI chatbot" that encompasses all AI chatbots, not just those marketed as romantic companions.
    - Rather than granting exemptions based on use cases (e.g. customer service chatbots), consider granting exemptions based on either (1) demonstrated safeguards and risk profiles through risk assessment process; or (2) structural platform restrictions on chatbot design features (e.g. chatbots that users cannot engage with for more than three hours continuously, or chatbots that users cannot re-engage with in ongoing, discontinuous conversations with retained memory)
- Require a static disclosure that chatbot is not human that is visible throughout a conversational session
- Address engagement-maximizing functionalities
    - Require periodic break reminders for all users every two hours
    - Require chatbots provide usage summaries at the end of each session and on account dashboard for users with accounts
    - Require friction-free exit (prohibit manipulative design features that make exiting a chat or deleting an account harder than staying) for all users
- Address intimate functionalities:
    - Provide a default model with limited sycophancy
    - Provide a default model that does not anthropomorphize (e.g. using personal pronouns, emulated empathy, etc.)

    - Provide a default model that is not capable of engaging in sexualized, romantic, or erotic conversations
    - Require account creation for users to access intimate functionalities / other model versions
- For minors:
    - Ensure there is an option to create an account and link an account to a parent account or dashboard
    - For minor account holders, require privacy settings to be set to the highest default (e.g. no training on chats, no data retention between chats)
    - For minor account holders, prohibit intimate functionalities by default
    - Provide clear, easy-to-use parent/guardian controls for minor accounts that allow parents to customize design features for minors in alignment with their developmental stage/needs
        - Parent/guardian controls should be opt-out rather than opt-in (i.e. default settings for minor accounts should be the most privacy-enhancing and design-constrained, with options to loosen these restrictions)

Principle 3: Mandate **transparency** and **accountability**

Policy options:

- Emotional and relational model risk assessment to be filed with DBHDS; DBHDS to lead consultative process with industry, VITA, VDH, and CAII to develop guidance for compliance
- Require providers to report adverse incidents to DBHDS and VDH (e.g. how many crisis referrals made, under what conditions, inappropriate model responses to crisis situations, etc.)
- Require providers to submit an annual report on user-submitted complaints to DBHDS; DBHDS to issue reporting guidance
- DBHDS to begin collecting data from community services boards on mental health crises associated with digital harms
- Attorney General's Office to impose civil penalties
- Empower the Attorney General's Office to investigate suspected violations of bill provisions

Note on oversight and enforcement: Virginia does not have a designated oversight agency with technical expertise to effectively implement provisions of an AI bill. JCOTS will continue ongoing discussions with the Executive branch to determine options for effective oversight.

Principle 4: Enable **user choice** and **autonomy**

Policy options:

- Require chatbot operators to offer clear, easy-to-use parent/guardian controls for minor accounts
- Require chatbot operators to provide clear, customizable options for all users to adjust design features, with default settings to restrict design features associated with these risks so users must proactively opt-in to use features like intimate functionality or follow-on questions

Principle 5: Enable **legislative learning**

Policy options:

- Ask JCOTS to report on data collected because of legislation (risk assessments, incident reports, user-submitted reports), evaluating the quality of reported information and insights gained from mandated reporting after 1 year, make recommendations for improvements

Principle 6: Engage **affected communities**

Policy options:

- Actively reach out to youth and youth organizations for input on draft legislation next session
- Ask JCOTS (perhaps in collaboration with VDH) for report on participatory methods for developing effective, autonomy-preserving online safety legislation targeted toward minors

---

Author Information
Kira Allmann, Ph.D.
Chief Policy Analyst, Joint Commission on Technology & Science
Contact: info@jcots.virginia.gov

Thanks To

Serena Oduro, *Data & Society*
Scott Babwah Brennen, *NYU Tech Policy Center*
Marissa Edmund, *Family Online Safety Institute (FOSI)*
Jennifer Gibbons, *Entertainment Software Association*
Tom Mann, *Computer & Communications Industry Association (CCIA)*
Julie Scelfo, *Mothers Against Media Addiction*
Adam Billen & Nathan Calvin, *Encode*
Zach Praiss, *RESET Tech*

For providing helpful information and background during the preparation of this report.

References

[1] Executive Order 14365, 90 Fed. Reg. 58499. 2025. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/

[2] Gottfried J et al. Americans and AI 2026: Chatbots, Smart Devices and Views on Impact. Pew Research Center; 2026. https://www.pewresearch.org/internet/2026/06/17/americans-and-ai-2026-chatbots-smart-devices-and-views-on-impact/

[3] Faverio M, Sidoti O. Teens, Social Media and AI Chatbots 2025. Pew Research Center. 2025 [accessed 2026 Apr 16]. https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/

[4] Valentino-DeVries J, Hill K. How Bad Are A.I. Delusions? We Asked People Treating Them. The New York Times. 2026 Jan 26 [accessed 2026 July 2]. https://www.nytimes.com/2026/01/26/us/chatgpt-delusions-psychosis.html

[5] Hill K. Lawsuits Blame ChatGPT for Suicides and Harmful Delusions. The New York Times. 2025 Nov 7 [accessed 2026 July 2]. https://www.nytimes.com/2025/11/06/technology/chatgpt-lawsuit-suicides-delusions.html

[6] Ipsos. Artificial Intelligence: Key insights, data and tables. 2026 June 29 [accessed 2026 July 21]. https://www.ipsos.com/en-us/artificial-intelligence-key-insights-data-and-tables

[7] Gottfried J et al. Americans and AI 2026: Chatbots, Smart Devices and Views on Impact. Pew Research Center; 2026. https://www.pewresearch.org/internet/2026/06/17/americans-and-ai-2026-chatbots-smart-devices-and-views-on-impact/

[8] AI Security Institute. Frontier AI Trends Report. AISI; 2025. https://www.aisi.gov.uk/frontier-ai-trends-report/pdf

[9] Zao-Sanders M. How People Are Really Using Gen AI in 2026. Harvard Business Review. 2026 June 1 [accessed 2026 June 11]. https://hbr.org/2026/06/how-people-are-really-using-ai-in-2026

[10] American Psychological Association. Patients are bringing AI to therapy. APA; 2026. https://www.apa.org/pubs/reports/chatbots-mental-health-2026

[11] Wang Z et al. History, Development, and Principles of Large Language Models-An Introductory Survey. 2024 [accessed 2025 Jan 21]. https://doi.org/10.48550/arXiv.2402.06853

[12] Isaac M. What the SpaceX I.P.O. Means for OpenAI and Anthropic. The New York Times. 2026 June 12 [accessed 2026 July 21]. https://www.nytimes.com/2026/06/12/technology/spacex-ipo-openai-anthropic.html

[13] Longpre S et al. Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them? MIT; 2024. https://mit-genai.pubpub.org/pub/uk7op8zs/release/2

[14] Ogilvie B. Scientific Archives in the Age of Digitization. Isis. 2016;107(1):77–85. https://doi.org/10.1086/686075

[15] The Centre for Internet and Society, Oxford Internet Institute, Whose Knowledge? State of the Internet's Languages Report. 2022. https://internetlanguages.org/en/summary/

[16] Sengupta A, Bouterse S, Allmann K. Build an internet for, and from, us all. Nature. 2018;563(7733):S147–S148. https://doi.org/10.1038/d41586-018-07506-7

[17] Guo E. A major AI training data set contains millions of examples of personal data. MIT Technology Review. 2025 July 18 [accessed 2025 July 21]. https://www.technologyreview.com/2025/07/18/1120466/a-major-ai-training-data-set-contains-millions-of-examples-of-personal-data/; Baack S. Training Data for the Price of a Sandwich: Common Crawl's Impact on Generative AI. Mozilla Foundation; 2024. https://foundation.mozilla.org/en/research/library/generative-ai-training-data/common-crawl/; Denton E et al. On the genealogy of machine learning datasets: A critical history of ImageNet. Big Data & Society. 2021;8(2):20539517211035955. https://doi.org/10.1177/20539517211035955

[18] Vecchione B, Singh R. Artificial intelligence is mental: Evaluating the role of large-language models in supporting mental health and well-being. Big Data & Society. 2025;12(4). https://doi.org/10.1177/20539517251383884

[19] Wang Z et al. History, Development, and Principles of Large Language Models-An Introductory Survey. 2024 [accessed 2025 Jan 21]. http://arxiv.org/abs/2402.06853. https://doi.org/10.48550/arXiv.2402.06853

[20] Hunter T. Chatbots are so gullible, they'll take directions from hackers. Washington Post. 2023 Nov 2 [accessed 2025 Feb 4]. https://www.washingtonpost.com/technology/2023/11/02/prompt-injection-ai-chatbot-vulnerability-jailbreak

[21] RoyChowdhury A et al. ConfusedPilot: Confused Deputy Risks in RAG-based LLMs. 2024 [accessed 2025 Feb 24]. http://arxiv.org/abs/2408.04870. https://doi.org/10.48550/arXiv.2408.04870

[22] Barman KG, Wood N, Pawlowski P. Beyond transparency and explainability: on the need for adequate and contextualized user guidelines for LLM use. Ethics and Information Technology. 2024;26(3):47. https://doi.org/10.1007/s10676-024-09778-2

[23] Wei M. Testimony of Marlynn Wei, M.D., J.D.: Hearing on "Innovation with Integrity: Examining the Risks and Benefits of AI Chatbots." 2025. https://democrats-energycommerce.house.gov/committee-activity/hearings/hearing-innovation-integrity-examining-risks-and-benefits-ai-chatbots-0; Wei M. Hidden Mental Health Dangers of Artificial Intelligence Chatbots. 2025 Oct 2 [accessed 2026 July 1]. https://www.psychologytoday.com/us/blog/urban-survival/202509/hidden-mental-health-dangers-of-artificial-intelligence-chatbots

[24] IEEE Society on Social Implications of Technology. IEEE Recommended Practice for Ethical Considerations of Emulated Empathy in Partner-Based General Purpose Artificial Intelligence Systems. The Institute of Electrical and Electronics Engineers, Inc.; 2026. https://standards.ieee.org/ieee/7014.1/11609/

[25] Suleyman M. AI is programmed to hijack human empathy — we must resist that. Nature. 2026;651(8106):559–559. https://doi.org/10.1038/d41586-026-00834-z

[26] Center for Humane Technology. Attachment Hacking and the Rise of AI Psychosis. Your Undivided Attention. 2026. https://www.humanetech.com/podcast/attachment-hacking-and-the-rise-of-ai-psychosis

[27] McStay A, Bakir V. Do AI Companions Understand? Most UK Teens Say Yes. Emotional AI Lab, Bangor University; 2026. https://andrewmcstay.org/publications/ai-companions-uk-teens/

[28] McStay A. Intimate Functionality — A Phrase Defining the Next Year of AI Regulation. Tech Policy Press. 2026 July 14 [accessed 2026 July 20]. https://techpolicy.press/intimate-functionality-a-phrase-defining-the-next-year-of-ai-regulation

[29] AI Security Institute. Frontier AI Trends Report. AISI; 2025. https://www.aisi.gov.uk/frontier-ai-trends-report/pdf

[30] IEEE Society on Social Implications of Technology. IEEE Recommended Practice for Ethical Considerations of Emulated Empathy in Partner-Based GeneralPurpose Artificial Intelligence Systems. The Institute of Electrical and Electronics Engineers, Inc.; 2026. https://standards.ieee.org/ieee/7014.1/11609/

[31] Hollanek T, Sobey A. AI Companions for Health and Mental Wellbeing: Opportunities, Risks and Policy Implications. 2025 Feb 7 [accessed 2025 Aug 18]. https://doi.org/10.17863/CAM.115939

[32] Faverio M, Sidoti O. Teens, Social Media and AI Chatbots 2025. Pew Research Center. 2025 [accessed 2026 Apr 16]. https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/; Gottfried J et al. Americans and AI 2026: Chatbots, Smart Devices and Views on Impact. Pew Research Center; 2026. https://www.pewresearch.org/internet/2026/06/17/americans-and-ai-2026-chatbots-smart-devices-and-views-on-impact/; Skjuve M, Brandtzaeg PB, Følstad A. Why do people use ChatGPT? Exploring user motivations for generative conversational AI. First Monday. 2024 Jan 3 [accessed 2026 July 23]. https://doi.org/10.5210/fm.v29i1.13541

[33] Zhang Y et al. The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. 2025 [accessed 2025 Aug 18]. http://arxiv.org/abs/2506.12605. https://doi.org/10.48550/arXiv.2506.12605; Ciriello R, Gal U, Turel O. Not a Silver Bullet for Loneliness: How Attachment and Age Shape Intimacy with AI Companions. 2026 [accessed 2026 Apr 16]. https://doi.org/10.48550/arXiv.2602.12476; Guingrich RE, Graziano MSA. A Longitudinal Randomized Control Study of Companion Chatbot Use: Anthropomorphism and Its Mediating Role on Social Impacts. 2025 [accessed 2026 June 30]. https://doi.org/10.48550/arXiv.2509.19515

[34] Faverio M, Sidoti O. Teens, Social Media and AI Chatbots 2025. Pew Research Center. 2025 [accessed 2026 Apr 16]. https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/; Gottfried J et al. Americans and AI 2026: Chatbots, Smart Devices and Views on Impact. Pew Research Center; 2026. https://www.pewresearch.org/internet/2026/06/17/americans-and-ai-2026-chatbots-smart-devices-and-views-on-impact/; Skjuve M, Brandtzaeg PB, Følstad A. Why do people use ChatGPT? Exploring user motivations for generative conversational AI. First Monday. 2024 Jan 3 [accessed 2026 July 23]. https://doi.org/10.5210/fm.v29i1.13541

[35] Bae Brandtzæg PB, Skjuve M, Kristoffer Dysthe KK, Følstad A. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2021. p 1–13. (CHI '21). https://doi.org/10.1145/3411764.3445318; Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB. My Chatbot Companion - a Study of Human-Chatbot Relationships. International Journal of Human-Computer Studies. 2021 [accessed 2025 Feb 11];149. https://doi.org/10.1016/j.ijhcs.2021.102601

[36] De Freitas J, Uğuralp AK, Uğuralp Z, Puntoni S. AI Companions Reduce Loneliness. 2024 [accessed 2025 Mar 17]. https://papers.ssrn.com/abstract=4893097. https://doi.org/10.2139/ssrn.4893097

[37] Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Mental Health. 2017;4(2). https://doi.org/10.2196/mental.7785

[38] Zhong W, Luo J, Zhang H. The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. Journal of Affective Disorders. 2024;356:459–469. https://doi.org/10.1016/j.jad.2024.04.057

[39] Ovsyannikova D, de Mello VO, Inzlicht M. Third-party evaluators perceive AI as more compassionate than expert humans. Communications Psychology. 2025;3(1):1–11. https://doi.org/10.1038/s44271-024-00182-6

[40] Stamatis C et al. Beyond Simulations: What 20,000 Real Conversations Reveal About Mental Health AI Safety. Research Square. 2026 Jan 27:rs.3.rs-8642399. https://doi.org/10.21203/rs.3.rs-8642399/v1

[41] Jafari K et al. Expert Evaluation and the Limits of Human Feedback in Mental Health AI Safety Testing. 2026 [accessed 2026 July 27]. http://arxiv.org/abs/2601.18061. https://doi.org/10.48550/arXiv.2601.18061

[42] See, for example: Zhang R et al. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025. p 1–17. http://arxiv.org/abs/2410.20130. https://doi.org/10.1145/3706598.3713429

[43] Vecchione B, Ye M, Garofalo L, Singh R. Engagement-Optimized Care: When LLMs become Mental Health Infrastructure. 2026 [accessed 2026 June 4]. https://arxiv.org/abs/2605.23787v1

[44] Ibrahim L et al. Sycophantic AI makes human interaction feel more effortful and less satisfying over time. 2026. https://arxiv.org/abs/2605.07912

[45] Liang K et al. Machine Bullshit: Characterizing the Emergent Disregard for Truth in Large Language Models. 2025 [accessed 2026 July 17]. https://doi.org/10.48550/arXiv.2507.07484; Nicholls L et al. "AI Psychosis" in Context: How Conversation History Shapes LLM Responses to Delusional Beliefs. 2026 [accessed 2026 July 2]. https://doi.org/10.48550/arXiv.2604.13860; Vecchione B, Ye M, Garofalo L, Singh R. Engagement-Optimized Care: When LLMs become Mental Health Infrastructure. 2026 [accessed 2026 June 4]. https://arxiv.org/abs/2605.23787v1

[46] Namvarpour M et al. Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2026. p 1–19. (CHI '26). https://doi.org/10.1145/3772318.3790597; Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/; Fraser HL, Ciriello R, Szczuka J. Regulating Artificial Intimacy: From Locks and Blocks to Relational Accountability. 2026 [accessed 2026 Apr 16]. https://doi.org/10.2139/ssrn.6078412; Vecchione B, Ye M, Garofalo L, Singh R. Engagement-Optimized Care: When LLMs become Mental Health Infrastructure. 2026 [accessed 2026 June 4]. https://arxiv.org/abs/2605.23787v1

[47] Namvarpour M et al. Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2026. p 1–19. (CHI '26). https://doi.org/10.1145/3772318.3790597; Ciriello R, Gal U, Turel O. Not a Silver Bullet for Loneliness: How Attachment and Age Shape Intimacy with AI Companions. 2026 [accessed 2026 Apr 16]. https://doi.org/10.48550/arXiv.2602.12476; Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/; Vecchione B, Ye M, Garofalo L, Singh R. Engagement-Optimized Care: When LLMs become Mental Health Infrastructure. 2026 [accessed 2026 June 4]. https://arxiv.org/abs/2605.23787v1; Ha T et al. How interactional AI may alter adolescent relational learning and mental health. The Lancet Child & Adolescent Health. 2026 [accessed 2026 July 2];0(0). https://doi.org/10.1016/S2352-4642(26)00166-5; Zhang R et al. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025. p 1–17. https://doi.org/10.1145/3706598.3713429

[48] Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/; Freitas JD, Oguz-Uguralp Z, Kaan-Uguralp A. Emotional Manipulation by AI Companions. 2025 [accessed 2026 July 17]. https://doi.org/10.48550/arXiv.2508.19258; Phang J et al. Investigating Affective Use and Emotional Well-being on ChatGPT. 2025 [accessed 2026 July 17]. https://doi.org/10.48550/arXiv.2504.03888

[49] Moore J et al. Characterizing Delusional Spirals through Human-LLM Chat Logs. 2026 [accessed 2026 June 23]. http://arxiv.org/abs/2603.16567.

https://doi.org/10.48550/arXiv.2603.16567; Zhang R et al. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025. p 1–17. https://doi.org/10.1145/3706598.3713429; Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[50] Zhang R et al. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025. p 1–17. https://doi.org/10.1145/3706598.3713429; 1. Jayathilake HM, Ma R. Anthropomorphism in Children's Interactions with LLM-Based Chatbots: A Systematic Review of Drivers and Outcomes. In: Proceedings of the 25th Annual ACM Interaction Design and Children Conference. Association for Computing Machinery; 2026. p 520–536. (IDC '26). https://dl.acm.org/doi/10.1145/3773077.3806126. https://doi.org/10.1145/3773077.3806126

[51] Vecchione B, Ye M, Garofalo L, Singh R. Engagement-Optimized Care: When LLMs become Mental Health Infrastructure. 2026 [accessed 2026 June 4]. https://arxiv.org/abs/2605.23787v1; Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[52] Gray CM. The Dark Patterns Knowledge Stack: Exploring New Ways to Negotiate Context, Law, and Design. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2026. p 1–11. (CHI '26). https://dl.acm.org/doi/10.1145/3772318.3791264. https://doi.org/10.1145/3772318.3791264; Chapman P, Steinberg M. Measuring Risk I: What EU Risk Assessments and US Litigation Reveal About Meta and TikTok. Knight-Georgetown Institute; 2026. https://kgi.georgetown.edu/research-and-commentary/measuring-risk-what-eu-risk-assessments-and-us-litigation-reveal-about-meta-and-tiktok/; Chapman P, Steinberg M. Measuring Risk II: What EU Risk Assessments and US Litigation Reveal About Snap and YouTube. Knight-Georgetown Institute; 2026. https://kgi.georgetown.edu/research-and-commentary/measuring-risk-what-eu-risk-assessments-and-us-litigation-reveal-about-snap-and-youtube/

[53] Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[54] Moore J et al. Characterizing Delusional Spirals through Human-LLM Chat Logs. 2026 [accessed 2026 June 23]. http://arxiv.org/abs/2603.16567. https://doi.org/10.48550/arXiv.2603.16567

[55] Fang CM et al. How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study. 2025 [accessed 2026 June 30]. http://arxiv.org/abs/2503.17473. https://doi.org/10.48550/arXiv.2503.17473

[56] Nicholls L et al. "AI Psychosis" in Context: How Conversation History Shapes LLM Responses to Delusional Beliefs. 2026 [accessed 2026 July 2]. https://doi.org/10.48550/arXiv.2604.13860

[57] Cheng M et al. Social Sycophancy: A Broader Understanding of LLM Sycophancy. 2025 [accessed 2025 June 2]. http://arxiv.org/abs/2505.13995. https://doi.org/10.48550/arXiv.2505.13995; Sharma M et al. Towards Understanding Sycophancy in Language Models. 2023 [accessed 2025 Jan 30]. http://arxiv.org/abs/2310.13548. https://doi.org/10.48550/arXiv.2310.13548

[58] Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[59] Metz C, Francisco KWM reported from S, Seattle KW from. A.I. Is Getting More Powerful, but Its Hallucinations Are Getting Worse. The New York Times. 2025 May 5 [accessed 2025 May 16]. https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html; Sun Y, Sheng D, Zhou Z, Wu Y. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. Humanities and Social Sciences Communications. 2024;11(1):1278. https://doi.org/10.1057/s41599-024-03811-x

[60] Brigham NG, Qin L, Kohno T. Examining Risks in the AI Companion Application Ecosystem. 2026 [accessed 2026 July 1]. https://arxiv.org/abs/2603.13620v1; Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[61] Airenti G. The Cognitive Bases of Anthropomorphism: From Relatedness to Empathy. International Journal of Social Robotics. 2015;7(1):117–127. https://doi.org/10.1007/s12369-014-0263-x; Guingrich RE, Graziano MSA. A Longitudinal Randomized Control Study of Companion Chatbot Use: Anthropomorphism and Its Mediating Role on Social Impacts. 2025 [accessed 2026 June 30]. http://arxiv.org/abs/2509.19515. https://doi.org/10.48550/arXiv.2509.19515; Jayathilake HM, Ma R. Anthropomorphism in Children's Interactions with LLM-Based Chatbots: A Systematic Review of Drivers and Outcomes. In: Proceedings of the 25th Annual ACM Interaction Design and Children Conference. Association for Computing Machinery; 2026. p 520–536. (IDC '26). https://dl.acm.org/doi/10.1145/3773077.3806126. https://doi.org/10.1145/3773077.3806126

[62] Freitas JD, Oguz-Uguralp Z, Kaan-Uguralp A. Emotional Manipulation by AI Companions. 2025 [accessed 2026 Feb 26]. http://arxiv.org/abs/2508.19258. https://doi.org/10.48550/arXiv.2508.19258; Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[63] IEEE Society on Social Implications of Technology. IEEE Recommended Practice for Ethical Considerations of Emulated Empathy in Partner-Based GeneralPurpose Artificial Intelligence Systems. The Institute of Electrical and Electronics Engineers, Inc.; 2026. https://standards.ieee.org/ieee/7014.1/11609/; Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[64] Ciriello R, Gal U, Turel O. Not a Silver Bullet for Loneliness: How Attachment and Age Shape Intimacy with AI Companions. 2026 [accessed 2026 Apr 16]. https://doi.org/10.48550/arXiv.2602.12476

[65] Ibrahim L, Hafner FS, Rocher L. Training language models to be warm can reduce accuracy and increase sycophancy. Nature. 2026;652(8112):1159–1165. https://doi.org/10.1038/s41586-026-10410-0

[66] Vecchione B, Ye M, Garofalo L, Singh R. Engagement-Optimized Care: When LLMs become Mental Health Infrastructure. 2026 [accessed 2026 June 4]. https://arxiv.org/abs/2605.23787v1

[67] Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[68] Mildner T et al. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2023. p 1–15. (CHI '23). https://doi.org/10.1145/3544548.3580695

[69] Zagal JP, Björk S, Lewis C. Dark Patterns in the Design of Games. 2013. https://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-24252

[70] Mathur A et al. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. Proceedings of the ACM on Human-Computer Interaction. 2019;3(CSCW):81:1-81:32. https://doi.org/10.1145/3359183

[71] Joshi R, Adjagbodjou A, Luria M. Dark Patterns in AI Chatbots: A Taxonomy to Inform Better Design. Center for Democracy & Technology; 2026. https://cdt.org/insights/dark-patterns-in-ai-chatbots-a-taxonomy-to-inform-better-design/

[72] OpenAI Model Spec. OpenAI. [accessed 2026 July 23]. https://model-spec.openai.com/2025-12-18.html#overview

[73] Claude's Character. Anthropic. [accessed 2026 July 23]. https://www.anthropic.com/research/claude-character

[74] Knight W. Microsoft's AI Boss Wants Copilot to Bring 'Emotional Support' to Windows and Office. Wired. [accessed 2025 Feb 11]. https://www.wired.com/story/mustafa-suleyman-interview-microsoft-ai-ceo-copilot/

[75] Roose K. A Conversation With Bing's Chatbot Left Me Deeply Unsettled. The New York Times. 2023 Feb 16 [accessed 2026 July 21]. https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html; See also: Crolic C, Thomaz F, Hadi R, Stephen AT. Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions. Journal of Marketing. 2022;86(1):132–148. https://doi.org/10.1177/00222429211045687; Larsen AG, Skjuve MB, Følstad A, van As N. LLM Hallucinations in Conversational AI for Customer Service: Framework and End-User Perceptions. International Journal of Human–Computer Interaction. 2026;42(13):9807–9828. https://doi.org/10.1080/10447318.2025.2580540

[76] Kuenssberg L. Mothers say AI chatbots encouraged their sons to kill themselves. BBC. 2025 Nov 8 [accessed 2026 July 23]. https://www.bbc.com/news/articles/ce3xgwyywe4o

[77] Young O. Colorado family sues AI chatbot company after daughter's suicide: "My child should be here." CBS Colorado. 2025 Oct 2 [accessed 2026 July 23]. https://www.cbsnews.com/colorado/news/lawsuit-characterai-chatbot-colorado-suicide/

[78] Chatterjee R. Their teenage sons died by suicide. Now, they are sounding an alarm about AI chatbots. NPR. 2025 Sept 19 [accessed 2026 July 16]. https://www.npr.org/sections/shots-health-news/2025/09/19/nx-s1-5545749/ai-chatbots-safety-openai-meta-characterai-teens-suicide

[79] Brewster RCL et al. Characteristics and Safety of Consumer Chatbots for Emergent Adolescent Health Concerns. JAMA network open. 2025;8(10). https://doi.org/10.1001/jamanetworkopen.2025.39022

[80] Fraser HL, Ciriello R, Szczuka J. Regulating Artificial Intimacy: From Locks and Blocks to Relational Accountability. 2026 [accessed 2026 Apr 16]. https://doi.org/10.2139/ssrn.6078412

[81] 1. Ragab A, Mannan M, Youssef A. "Trust Me Over My Privacy Policy": Privacy Discrepancies in Romantic AI Chatbot Apps. In: 2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). 2024. p 484–495. https://ieeexplore.ieee.org/abstract/document/10628933. https://doi.org/10.1109/EuroSPW61312.2024.00060

[82] Brigham NG, Qin L, Kohno T. Examining Risks in the AI Companion Application Ecosystem. 2026 [accessed 2026 July 1]. https://arxiv.org/abs/2603.13620v1

[83] Common Sense Media. Talk, Trust, and Trade-offs: How and Why Teens Use AI Companions. 2025. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

[84] McBain RK et al. Use of Generative AI for Mental Health Advice Among US Adolescents and Young Adults. JAMA Network Open. 2025 [accessed 2026 July 2];8. https://www.rand.org/pubs/external_publications/EP71039.html

[85] Commission on Youth. Collection of Evidence-based Practices for Children and Adolescents with Mental Health Treatment Needs - 10th Edition. Commonwealth of Virginia; 2025. https://vcoy.virginia.gov/collection.asp

[86] The U.S. Surgeon General's Advisory. Social Media and Youth Mental Health. U.S. Department of Health and Human Services; 2023. https://www.hhs.gov/surgeongeneral/reports-and-publications/youth-mental-health/social-media/index.html; Commission on Youth. Collection of Evidence-based Practices for Children and Adolescents with Mental Health Treatment Needs - 10th Edition. Commonwealth of Virginia; 2025. https://vcoy.virginia.gov/collection.asp

[87] Orben A. Teenagers, screens and social media: a narrative review of reviews and key studies. Social Psychiatry and Psychiatric Epidemiology. 2020;55(4):407–414. https://doi.org/10.1007/s00127-019-01825-4

[88] Written Evidence submitted by the Digital Mental Health Group. Reference No. SMR0005. 2026. https://committees.parliament.uk/writtenevidence/163429/default/

[89] Brewster RCL et al. Characteristics and Safety of Consumer Chatbots for Emergent Adolescent Health Concerns. JAMA network open. 2025;8(10):e2539022. https://doi.org/10.1001/jamanetworkopen.2025.39022

[90] Ha T et al. How interactional AI may alter adolescent relational learning and mental health. The Lancet Child & Adolescent Health. 2026 [accessed 2026 July 2];0(0). https://doi.org/10.1016/S2352-4642(26)00166-5

[91] Namvarpour M et al. Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2026. p 1–19. (CHI '26). https://doi.org/10.1145/3772318.3790597

[92] Skjuve M, Følstad A, Brandtzaeg PB. ChatGPT as a mental health advisory service: Comparing evaluations from youth and health professionals. DIGITAL HEALTH. 2026 [accessed 2026 May 19];12. https://doi.org/10.1177/20552076261427447

[93] Namvarpour M et al. Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2026. p 1–19. (CHI '26). https://doi.org/10.1145/3772318.3790597; 1. Ha T et al. How interactional AI may alter adolescent relational learning and mental health. The Lancet Child & Adolescent Health. 2026 [accessed 2026 July 2];0(0). https://doi.org/10.1016/S2352-4642(26)00166-5

[94] Common Sense Media. Talk, Trust, and Trade-offs: How and Why Teens Use AI Companions. 2025. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

[95] Livingstone S, Sylwander KR. There is no right age! The search for age-appropriate ways to support children's digital lives and rights. Journal of Children and Media. 2025;19(1):6–12. https://doi.org/10.1080/17482798.2024.2435015

[96] Health advisory: Artificial intelligence and adolescent well-being. The American Psychological Association. [accessed 2026 July 2]. https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being

[97] boyd d. it's complicated. Yale University Press; 2014; Masur PK, Veldhuis J, Vaate NB de. There Is No Easy Answer: How the Interaction of Content, Situation, and Person Shapes the Effects of Social Media Use on Well-Being. In: The Social Media Debate. Routledge; 2022.

[98] Livingstone S et al. Children's Rights and Online Age Assurance Systems. 2024 Oct 24 [accessed 2026 July 28]. https://brill.com/view/journals/chil/32/3/article-p721_009.xml. https://doi.org/10.1163/15718182-32030001

[99] Livingstone S et al. Children's Rights and Online Age Assurance Systems. 2024 Oct 24 [accessed 2026 July

28]. https://brill.com/view/journals/chil/32/3/article-p721_009.xml. https://doi.org/10.1163/15718182-32030001; Wodo W, Gorski M, Hanzlik L. Age Verification in the Web - Holy Grail to Control Access to Restricted Content. 2026 [accessed 2026 July 9]. http://arxiv.org/abs/2605.05513. https://doi.org/10.48550/arXiv.2605.05513

[100] Minocha S et al. Papers, Please: A First Look at Age Verification on the Web. 2026. https://mikespecter.com/assets/pdf/AgeVerification.pdf; Rescorla E, Arnao Z, Cooper A. Age Assurance Online: A Technical Assessment of Current Systems and Their Limitations. Knight-Georgetown Institute; 2026. https://kgi.georgetown.edu/wp-content/uploads/2026/01/Age_Assurance_Online_Technical-Assessment_Report_KGI.pdf

[101] IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children. IEEE Std 2089-2021. 2021. https://doi.org/10.1109/IEEESTD.2021.9627644

[102] Age-Appropriate Design Codes. Kids Code Coalition. 2026 [accessed 2026 July 28]. https://kidscodecoalition.org/age-appropriate-design-codes/

[103] Reid A, Neill L, Ringel E. Nerd Harder: A Typology of Techno-Legal Solutionist Logics in Child Online Safety Laws. Policy & Internet. 2025;17(3):e70012. https://doi.org/10.1002/poi3.70012

[104] Health advisory: Artificial intelligence and adolescent well-being. The American Psychological Association. [accessed 2026 July 2]. https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being

[105] Building towards age prediction. OpenAI. 2026 July 23 [accessed 2026 July 24]. https://openai.com/index/building-towards-age-prediction/

[106] Guide your child's Gemini Apps experience - Google For Families Help. Google For Families Help. [accessed 2026 July 24]. https://support.google.com/families/answer/16109150?hl=en#added_protections

[107] Our Approach to Teen AI Safety: Empowering Parents, Protecting Teens. Meta Newsroom. 2025 Oct 17 [accessed 2026 July 24]. https://about.fb.com/news/2025/10/teen-ai-safety-approach/

[108] Malik A. Snapchat now lets parents restrict their teens from using the app's 'My AI' chatbot. TechCrunch. 2024 Jan 11 [accessed 2026 July 24]. https://techcrunch.com/2024/01/11/snapchat-lets-parents-restrict-their-teens-using-apps-my-ai-chatbot-view-their-privacy-settings/

[109] Taking Bold Steps to Keep Teen Users Safe on Character.AI. CharacterAI blog. 2025 Oct 29 [accessed 2026 July 24]. https://blog.character.ai/u18-chat-announcement/

[110] Independent reviews of the AI used by kids. Common Sense Media. [accessed 2026 July 27]. https://institute.commonsensemedia.org/risk-assessments

[111] Wang J, Huang K, Klyman K, Bommasani R. Do AI Companies Make Good on Voluntary Commitments to the White House? 2025 [accessed 2026 June 1]. https://doi.org/10.48550/arXiv.2508.08345; Béjar A. Teen Accounts Broken Promises: How Instagram is failing to protect minors. Fairplay; 2025. https://fairplayforkids.org/wp-content/uploads/2025/09/Teen-Accounts-Broken-Promises-How-Instagram-is-failing-to-protect-minors.pdf

[112] Orben A, Matias JN. Fixing the science of digital technology harms. Science. 2025;388(6743):152–155. https://doi.org/10.1126/science.adt6807

[113] Béjar A. Teen Accounts Broken Promises: How Instagram is failing to protect minors. Fairplay; 2025. https://fairplayforkids.org/wp-content/uploads/2025/09/Teen-Accounts-Broken-Promises-How-Instagram-is-failing-to-protect-minors.pdf; Valentino-DeVries J. Many Child Safety Features on Social Apps Don't Work, Report Finds. The New York Times. 2026 June 29 [accessed 2026 June 30].

https://www.nytimes.com/2026/06/29/us/social-media-child-safety-report.html

[114] Haugen F. Statement of Frances Haugen. U.S. Senate Committee on Commerce, Science and Transportation, Sub-Committee on Consumer Protection, Product Safety, and Data Security. 2021. https://www.commerce.senate.gov/wp-content/uploads/media/doc/Frances%20Haugen%20Written%20Testimony.pdf

[115] Valentino-DeVries J. 'Teachers Are Going to Hate It': How Social Media Apps Hooked Teens at School. The New York Times. 2026 June 4 [accessed 2026 July 20]. https://www.nytimes.com/2026/06/04/us/social-media-schools.html

[116] New Mexico Department of Justice. "New Mexico Department of Justice Wins Landmark Verdict Against Meta." March 24, 2026. https://nmdoj.gov/press-release/new-mexico-department-of-justice-wins-landmark-verdict-against-meta/.

[117] Huamani K, Ortutay B. Instagram and YouTube found liable in landmark social media addiction trial in California. PBS News. 2026 Mar 25 [accessed 2026 July 20]. https://www.pbs.org/newshour/nation/instagram-and-youtube-found-liable-in-landmark-social-media-addiction-trial-in-california

[118] Valentino-DeVries J. 'Teachers Are Going to Hate It': How Social Media Apps Hooked Teens at School. The New York Times. 2026 June 4 [accessed 2026 July 20]. https://www.nytimes.com/2026/06/04/us/social-media-schools.html

[119] Orben A. Teenagers, screens and social media: a narrative review of reviews and key studies. Social Psychiatry and Psychiatric Epidemiology. 2020;55(4):407–414. https://doi.org/10.1007/s00127-019-01825-4

[120] Future of Privacy Forum. 2026 Chatbot Legislation Tracker. https://fpf.org/. [accessed 2026 July 21]. https://fpf.org/2026-chatbot-legislation-tracker/

[121] Lima-Strong C. Inside the Lobbying Frenzy Over California's AI Companion Bills. Tech Policy Press. 2025 Sept 23 [accessed 2026 July 24]. https://techpolicy.press/inside-the-lobbying-frenzy-over-californias-ai-companion-bills

[122] 1. Gluck J. Understanding the New Wave of Chatbot Legislation: California SB 243 and Beyond - Future of Privacy Forum. Future of Privacy Forum. 2025 [accessed 2026 July 24]. https://fpf.org/blog/understanding-the-new-wave-of-chatbot-legislation-california-sb-243-and-beyond/

[123] Mak A. AI's other free speech fight. POLITICO. 2026 July 23 [accessed 2026 July 23]. https://www.politico.com/newsletters/digital-future-daily/2026/06/29/ais-other-free-speech-fight-00980377

[124] Kang C, McCabe D. State Attorneys General Are Investigating OpenAI. The New York Times. 2026 June 13 [accessed 2026 July 20]. https://www.nytimes.com/2026/06/13/technology/states-investigating-openai.html

[125] Shapiro Administration Sues Character.AI Over Fake Medical Claims. Commonwealth of Pennsylvania. 2026 May 5 [accessed 2026 July 23]. https://www.pa.gov/governor/newsroom/2026-press-releases/shapiro-administration-sues-character-ai-over-fake-medical-claim

[126] AG Coleman Sues AI Chatbot Company for Preying on Children. Kentucky.gov. 2026 Jan 8 [accessed 2026 July 23]. https://www.kentucky.gov/Pages/Activity-stream.aspx?n=AttorneyGeneral&prId=1857

[127] FTC Launches Inquiry into AI Chatbots Acting as Companions. Federal Trade Commission. 2025 Sept 11 [accessed 2026 July 1]. https://www.ftc.gov/news-events/news/press-releases/2025/09/ftc-launches-inquiry-ai-chatbots-acting-companions

[128] Attorney General Ken Paxton Investigates Meta and Character.AI for Misleading Children with Deceptive AI-Generated Mental Health Services | Office of the Attorney General. Texas Office of the Attorney General. 2025 Aug 18 [accessed 2026 July 23]. https://www.texasattorneygeneral.gov/news/releases/attorney-general-ken-paxton-investigates-meta-and-characterai-misleading-children-deceptive-ai

[129] Bipartisan Coalition of State Attorneys General Issues Letter to AI Industry Leaders on Child Safety. National Association of Attorneys General. 2025 Aug 26 [accessed 2026 July 23]. https://www.naag.org/press-

releases/bipartisan-coalition-of-state-attorneys-general-issues-letter-to-ai-industry-leaders-on-child-safety/

[130] Growing Up in a Digital World 2030: Digital Childhoods. The Lancet & Financial Times Commission; 2021. https://www.governinghealthfutures2030.org/policy-brief-digital-childhoods/

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Bill Study

App Store Accountability Act

August 2026

How to read this report

This bill study is organized into nine sections as follows:

- The **Executive Summary** provides a synopsis of the bill study and presents the key takeaways;
- **Key Concepts** provides definitions for important terms used throughout the report;
- The **Introduction** explains what is the problem addressed in the bill, what evidence supports the problem, and why it matters to the Commonwealth;
- **Understanding Age Assurance** provides a technical overview of the methods for assessing the age of users and the tradeoffs in accuracy, privacy and accessibility;
- The **Policy Landscape** compares Virginia's introduced bill to legislation in other states and reports what is known about their outcomes;
- Finally the **Policy Recommendation** section puts forth several recommendations for policymakers to consider.

The report can be read as a continuous document, or be skipped to sections of interest. The main compelling ideas are presented in the Policy Recommendation section.

Executive Summary

This study reviews the evidence surrounding age assurance and considers the concerns raised by placing age verification and parental consent duties at the app store level. Available data supports the concerns that prompted SB237, but it does not show how often these harms occur or how effectively app store age verification would prevent them. Furthermore, similar laws in other states have been enacted within a short time frame and have yet to produce measurable results. In addition, recent age assurance legislation in other states has been subject to litigation, substantial amendments, and delays in implementation.

The study presents five recommendations for policymakers to consider. First, to improve the quality and reliability of information users have about apps, an independent body should audit age ratings and content descriptions assigned by app developers. Second, to decrease the risk of consent fatigue, parental controls should be easy to understand and use, and controls should be set to the highest default safeguards with options to remove safeguards. Third, to improve the information policy makers and technologists have about risks and harms, users should have a clear way to report adverse experiences in app stores. Fourth, parental consent should apply only to apps whose age rating exceeds

the user's age category, or to apps that involve purchases. Finally, to enhance accountability, app store providers should collect and publicly report the data needed to assess how these requirements work in practice.

Key Concepts

Age assurance: Refers to the overarching methods employed to establish, verify or estimate the age or age range of an individual. These methods include age verification, age estimation, and age inference; a fourth approach, age attestation, involves no confirmation and is treated as a weak control. Furthermore, these methods provide different degrees of certainty and may have different implications for privacy, costs, accuracy, and accessibility [1, 2]. Age assurance does not necessarily require confirmation of identity: age can be treated as a separate attribute from identity

Age attestation: A declaration of age made by a user, or by another person on the user's behalf, without a direct confirmation from official identification documents. Attestation is not among the three core methods recognized in the international standard on age assurance [1], which treats an unchecked declaration as a weak control. Nonetheless, app store accountability acts in the US allow an affirmative attestation by a parent or legal guardian of a minor as a method for satisfying the requirements in age-category verification [3-6].

Age estimation: An age assurance method that analyzes biological or behavioral attributes associated with age to determine the age or age range. Facial analysis is one common example of this method. Since this estimation produces a likelihood rather than a factual confirmation, errors may occur near the legal age thresholds that determine eligibility [1, 7].

Age inference: An age assurance method that uses verified information to draw an indirect conclusion about the age or age range of an individual. Such information may include for instance, the ownership of a financial product that is available only to adults [1].

Age rating: One or more classifications that indicate the suitability of the content and functions on an app or in-app purchase, to the user associated with an age category. Under app store accountability acts, the developer is responsible for assigning the age ratings. However, unlike other types of ratings (e.g., films), app ratings are not required to be issued or validated independently.

Age signal: Information related to an age category or age range that an operating system or app store transmits to a developer. Developers may use the signal to apply age-related restrictions and protective settings. An age signal communicates information about age rather than a decision made by a parent concerning a particular app. Major technology platforms have begun providing age signal interfaces, although the access is generally limited to jurisdictions where legislation requires the service [8, 9].

Age verification: An age assurance method that calculates the age of a user from a date of birth confirmed from an official source (e.g., passport or driver's licence). Age verification offers the strongest assurance but it also excludes users who do not possess the required documents [1].

App store: An online resource that is publicly available (i.e., website, software application, or other electronic service or platform) and allows account holders to download and install apps created by third-party developers.

Content description: A description of the specific elements or functions that inform an app's age rating. In general, the developer is required to provide both the app content description and the age rating.

Data minimization: The principle that an organization should collect only personal data that is adequate and reasonably necessary for a disclosed intention. Data minimization is a core requirement of the Virginia Consumer Data Protection Act [10, 11].

Significant change: A modification to the terms of service or privacy policy of an app that changes the categories of data collected, stored, or shared. The term also includes changes to the age rating or content description, and applies when an app adds in-app purchases or advertising that it did not previously include. The term was first used in Utah's App Store Accountability Act from 2025 and was adopted into the state bills that followed. The closest analogue in federal law is the "material change" under COPPA.

Verifiable parental consent: An authorization granted by a parent or legal guardian before a specified action involving a minor may proceed. Under COPPA, the term refers to consent for the collection, use, or disclosure of personal information from a child younger than 13 [12].

Zero-knowledge proof (ZKP): A cryptographic method that allows an individual to prove a claim, such as being older than 18, without disclosing an underlying identity document or other personal information. ZKP-based age assurance has been deployed by the European Commission and by Google Wallet [13, 14].

Introduction

For the last decade, mobile applications ("apps") have been the primary way for young people to engage with the digital world, and app stores have been the most popular means of accessing those apps. An app store allows an individual to download mobile apps created by third-party developers onto a mobile device. Apple and Google are the two largest app store providers, although other companies also operate app stores. Most of the mobile apps that are used by young people today, including games, social media, text messaging services, and chatbots, can be downloaded through an app store. A 2024 Pew Research Center survey found that 95% of U.S. teenagers ages 13 to 17 had access to a smartphone at home, and nearly half said they were online almost constantly [15]. The US Surgeon General

similarly reported that approximately 95% of adolescents use at least one social media platform, with more than one-third using social media almost all the time [16].

There has been an increase in public concern regarding the amount of time children spend using digital media and the specific content they engage with. The Centers for Disease Control and Prevention (CDC) in its 2023 Youth Risk Behaviour Survey found that 16% of US high school students experienced cyberbullying during the previous year [17]. Additionally, the Surgeon General reported that approximately two-thirds of adolescents across the US were often or sometimes exposed to hate-based content on social media [16]. A separate study by Common Sense Media found that 73% of teenagers between 13 and 17 had seen pornography on the internet. It is noteworthy that among all respondents, 29% reported having encountered that content only by accident, without ever intentionally searching for it [18]. A separate report by the same organization found that nearly three in four teenagers had used AI companion applications, with half using them regularly, and that around one-third of those users had felt uncomfortable with something an AI companion said or did [19]. The Federal Trade Commission has since opened an inquiry into how providers of such applications evaluate their safety for children and teenagers, although that study is not an enforcement action and has not yet been reported [20]. App stores do not directly cause these harms because harmful conduct occurs after accessing an app. However, app stores constitute an important gatekeeper in accessing the content consumers are concerned about. Therefore, app stores may be an important site for intervention before a minor engages with harmful content.

App store users have historically had limited information available before deciding whether to download an app. In a 2012 review of children's mobile apps, the Federal Trade Commission found that disclosures about data collection and in-app purchases were often missing before installation. Only 20% of the apps provided a privacy disclosure before download, and only 15% disclosed advertising, even though 58% contained it [21]. In addition, the FTC found that the information provided after download could be too late for parents or guardians to take action, as children may have already been using the app or a parent may have already been charged [22]. Although these findings are over ten years old, more recent research suggests that the gap has not been closed. A 2023 study of 20,195 Google Play apps for children found that approximately 81% of those designed exclusively for children used trackers, and that around 19% reported age ratings that were inconsistent across rating authorities [23]. Additionally, a 2026 study of nearly six thousand child-accessible mobile apps found that while both major app stores (i.e., Apple and Google) provide privacy labels for mobile apps, there was a high level of inconsistency between the labels an app displayed on each store, indicating that these disclosures do not reliably reflect what an app actually collects and shares [24].

One of the most clearly documented concerns is financial harm. For example, in 2014, as part of a settlement agreement with the FTC, Apple agreed to issue at least $32.5 million in refunds to parents who were charged for purchases made by their children without their consent [25]. As of June 2025, the FTC distributed more than $126 million to Fortnite players as a result of complaints involving unauthorized purchases of in-game items made using the credit cards of parents without their knowledge [26].

App Store Accountability in the News

- "Supreme Court won't block Texas app store age verification law," *Reuters*, 6 July 2026
- "Louisiana sues Roblox alleging the popular gaming site fails to protect children," *Associated Press*, 14 August 2025
- "Utah becomes first state to pass legislation requiring app stores to verify user ages and get parental consent for minors," *Associated Press*, 6 March 2025
- "App stores are set to verify ages under a new Utah law — and Meta, Snap, and X are thrilled," *Business Insider*, 6 March 2025
- "FTC Sends $126 Million in Refunds to Fortnite Players Who Were Charged for Unwanted Items, Reopens Claims Process," *Federal Trade Commission*, 25 June 2025
- "France Joins a Global Move Toward Restricting Social Media for Children," The New York Times, 21 July 2026

The growing public concern on these issues has begun to translate into state law, as existing federal protections have proven inadequate for addressing the policy issues around apps. The Children's Online Privacy Protection Act (COPPA) was designed to regulate the collection of personal data for children under 13, and therefore leaves younger and older minors outside its protection [12]. Furthermore, the parental consent required by COPPA covers only the management of personal information and not the action of downloading or making app-related purchases. COPPA has many well-evidenced weaknesses in the current digital environment. The FTC released a report in 2024 detailing how major online services often treated teens no differently from adults when collecting their data, covering practices from 2019 and 2020 [27]. The report noted that most online services do not place any restrictions on minors accessing the services they offer, and it recommended that Congress extend privacy protections to minors over 13. However, Congress has not enacted such protections yet. In the absence of federal action, several states have begun to introduce their own laws to regulate the access of minors to mobile apps.

SB237 enters this developing state policy landscape with a basic premise: the tools parents need already exist, but providers do not offer them consistently, and there are no mechanisms to hold providers accountable. For example, Apple, Google, Samsung, and Amazon each already offer some capabilities for capturing age signals or requesting parental consent. However, each company makes these tools available only in the jurisdictions where legislation requires them, rather than across all markets [8, 9, 28, 29]. In some cases, providers have publicly opposed the legislation requiring these tools. For example, Google has raised concerns with age verification laws from the Texas App Store Accountability Act (SB2420)[5] in their developer manuals [30]. Similarly, Apple highlighted that those laws would require the collection of sensitive, personally identifiable information to download any app, "even if a user simply wants to check the weather or sports scores" [31].

Additionally, some evidence suggests that parents want more choices about how and when children download apps, but they also find managing children's device use to be challenging. In a survey conducted by Pew in late 2023, it was reported that 76% of parents considered managing their children's phone time an important or top priority, while only 47% reported that they already limit it. Additionally, 43% of parents described the task as hard, compared to 26% who found it easy [32]. In short, proposed legislation, including SB237, aims to address the challenges parents and minors are encountering when trying to exercise choice over what content they want to access. It would require app stores to provide, before every download or purchase, information that users need to determine whether to download a specific app and options for parental oversight that parents desire but currently lack in most jurisdictions.

The available evidence supports the existence of the problem that app-store age-verification aims to solve, but it does not demonstrate whether it would successfully address those concerns. Therefore, three questions remain open: 1) whether better disclosure and a parental consent requirement actually lead parents to make better decisions; 2) whether the app store is the right point of intervention, given what it can and cannot reach; and 3) whether the age of every account holder can be assessed without creating the privacy risks that the bill seeks to prevent. The sections that follow address each of these questions.

Understanding Age Assurance

In order to safeguard minors online, providers rely on a family of methods collectively referred to as age assurance, which differ substantially in how they work and what their tradeoffs are in terms of privacy, user burden, and cost. Different methods of age assurance may determine how well a legislative requirement works in practices. The international standard on age assurance (i.e., ISO/IEC 27566-1) recognizes three main methods of age assurance: verification, estimation, and inference [1].

Age verification checks a user's age against an authoritative source (e.g., passport, driver's licence, credit card, etc.). Because it relies on a verified record rather than an estimate, it provides the strongest assurance of the three methods. That higher level of certainty comes with a greater burden. Users must have and submit an official document, which may exclude people who do not possess one, and the provider must collect identity information that some users may be reluctant to share [1]. **Age estimation** does not rely on an official document. Instead, it uses physical or behavioral attributes, most commonly facial analysis, to estimate a person's age or age range. This approach may be less intrusive, but it cannot provide the same degree of certainty, particularly near legal thresholds such as the boundary between ages 16 and 18 [7]. **Age inference** uses other verified information to draw an indirect conclusion about age. For example, ownership of a financial product available only to adults may indicate that a user is over 18. The reliability of this method depends on how closely the underlying information is connected to age; a weak or indirect signal provides limited assurance.

An important concern is that age estimation, and to a greater extent age verification, may require an app store to process personal information or biometric data for every user, not only for minors. Therefore, in order to identify which users are minors, an app store provider may first assess the age for every account, effectively creating an age-gate for all users. The selected method used by an app store provider will largely determine how much data must be collected and to whom the method applies.

These methods also differ in who performs the check. An app store may assess age itself, or it may delegate the task to a specialist age verification provider. Delegation carries its own tradeoff, since it would imply sharing the user's personal information with third-parties [33]. In addition, a technical assessment published by the Knight-Georgetown Institute found that no single method is sufficient on its own because self-declaration, commercial and government records, government identification documents, and age estimation all have limitations in either precision or availability [33]. Thus, the report suggests that a good age assurance system must support several methods, so that the user can use alternatives in case one of these is not applicable in a specific situation.

Many legislative requirements to safeguard minors do not specify the technical means of age assurance, leaving this determination up to providers. This ambiguity introduces additional risks: mandating age assurance may give providers a blank slate to collect even more data about all users, or it may introduce new security risks when third party age verifiers are enlisted to conduct age assurance.

Applying age assurance in app stores

The question for policymakers is whether age assurance, when applied at the app store, would protect children from the harms they encounter through apps. Answering this question requires two

conditions to hold. First, the app store must be able to establish the age category of its account holders. Second, the apps themselves must have accurate information about their content, so that an age category can be matched to an appropriate restriction. Unfortunately, currently neither condition is fully satisfied.

Placing age assurance at the app store carries a structural advantage that allows the assessment to be performed once at a single point instead of multiple times by every developer. Thus, centralizing age assurance would reduce the burden on users and the number of parties that must hold personal data. However, because age verification is associated with an individual account, it may not always be accurate in reflecting the actual age of the user operating the device at that particular time, especially if several users share a device.

The second condition concerns whether the information presented to the account holder truly reflects the app's actual content. Each app is assigned an age rating based on a questionnaire completed by the developer, without independent confirmation. On Google Play, these responses are based on the International Age Rating Coalition (i.e., IARC) system, and although member authorities may review and override a rating, they generally do so through spot checks and user complaints rather than through pre-release inspection [34, 35]. Apple, on the other hand, assigns ratings through its own questionnaire, without a coalition of rating authorities [36]. Moreover, a five-year study of the Google Play Store found that hundreds of top apps carried content ratings that did not initially match their actual content and were corrected only later. Moreover, ratings can change without notifying the user so a child who downloaded an app under one rating may afterward be exposed to content under the revised rating [37]. Furthermore, because a higher rating would result in a smaller audience, developers face a financial incentive to provide a conservative, lower self-rating in the questionnaire.

Treating the app stores as the centralized point through which minors reach apps could be classified simultaneously as *too broad* and *too narrow*. It is *too broad* in that it applies to all apps rather than only to the categories that are the most associated with online harm (e.g., social media, messaging, or content-sharing apps), extending the same requirements to apps that carry little or no relevant risk (e.g., bookstores, learning apps). There is often a misconception that what the bill proposes would be the equivalent of requesting to show an identification for a restricted item in a physical store (e.g., alcohol), however this is not a valid analogy in this context, since a store checks identification only at the moment a restricted product is purchased, whereas app-store age verification would require every user to have their age assessed simply to enter the app store, and would then require parental consent for a minor to download any app at all [38, 39]. Therefore, the requirement would arguably reach well beyond the categories of apps that motivated the bill: the age-assessment burden falls on every user, and the consent burden falls on minors across all apps rather than only those carrying age-restricted content

[39]. The technical literature supports this distinction by identifying two separate use cases for age assurance: applying "safer defaults" for general-purpose applications versus restricting access to specific content categories [33]. Of course, without accurate, independently audited age ratings for apps, there is no way to determine with certainty which apps carry higher risks for minors.

At the same time, it's *too narrow* because minors can often access the same content outside of app stores. Although the bill's definition of "app" is broad enough to include a web browser itself, Virginia's proposed bill would regulate the download of apps through an app store and likely would not apply to the websites and web-based services a minor visits through a browser once it is installed. For example, a minor could download a browser, such as Google Chrome, and navigate to Instagram using a URL and log into an Instagram account through the browser—all without downloading the Instagram app. As all modern devices ship with a web browser installed, a minor would still have the ability to access a service via an Internet browser without having access to the app store or going through the parental consent process [40].

There are also additional gaps that arise at the point of distribution for mobile apps; because the bill's definition of "app store" refers to platforms that host apps from third-party developers, a developer that distributes only its own app from its own website falls outside both the "app store" and the "developer" definitions [40, 41], and developers using payment methods outside the purchasing system of an app store must implement their own parental consent mechanisms [29]. Additional evidence suggests that when the coverage of regulations is inconsistent, users would be inclined to use the least-regulated services. For example, after Louisiana required age verification for adult websites, search interest decreased for compliant sites and increased for non-compliant sites [42]. As such, while age assurance through the major centralized app stores would encompass many of the avenues users take to download apps, it does not address web access and first-party distribution models.

Challenges to Age Assurance

There are several important challenges and critiques of age assurance, including limits of effectiveness, risks to privacy and user rights, and the implications of layering parental consent on top of age assurance mechanisms. The Knight-Georgetown Institute has evaluated age assurance systems by considering four main criteria: (i) baseline accuracy, (ii) resistance to circumvention, (iii) availability, and (iv) privacy [33]. In so doing, they found that no one mechanism works well across all four dimensions, and that these trade-offs are inherent for all current age assurance approaches, not a result of immature technology. The two subsections that follow groups the criteria into two key questions: whether age assurance is effective and what it costs in terms of privacy and user rights.

A) Effectiveness

The most comprehensive assessment of these methods to date is the Australian Age Assurance Technology Trial, an independent evaluation of 48 providers and more than 60 technologies, conducted before the introduction of minimum age requirements for social media [43-45]. The trial concluded that age assurance "can be done in Australia privately, efficiently and effectively," and that there are no substantial technological barriers preventing its deployment. However, the trial did also specify that no single solution would suit all the use cases, and therefore, no specific method can guarantee to be effective in all circumstances.

To date, Australia has produced the most compelling evidence of how such requirements perform in practice. Researchers at the University of Newcastle surveyed adolescents before and three months after the Social Media Minimum Age Act took effect in December 2025, and found that more than 85% of respondents under 16 continued to access the restricted platforms, even though two-thirds of them reported encountering an age check [46]. The form of verification they encountered the most was a self-declaration of age, although the eSafety Commissioner indicated that such a method was insufficient not long before the Act took effect [72]. Between 54% and 68% continued using the accounts they had before the regulation took place, while a smaller group used fake accounts, accessed services through other people's accounts, or relied on private browsing [46]. Surprisingly, a separate survey conducted on children that still used restricted platforms found that only 4% to 5% had used a VPN or a comparable tool, and that most had not needed a workaround at all, because the platforms had never identified their accounts in the first place [47]. The authors attributed this outcome not to the ingenuity of minors but to age checks that were not effective.

Circumvention is often assumed to be the principal obstacle to age assurance, but the available evidence gives limited support to that assumption. After the United Kingdom introduced the Online Safety Act in July 2025, several VPN providers reported sharp increases on sign-ups, in some cases of more than 1,000% [48]. Although this suggests an intention to bypass age checks, it does not establish that circumvention actually occurred. Furthermore, the evidence from Australia indicates that technical circumvention occurs in a small proportion of minors who continue to access restricted platforms. A qualitative study of fifteen Australian children aged 12 to 16 found that where circumvention did occur, the most common and lowest-effort method was simply falsifying data, by entering a false date of birth or using an account belonging to an older relative or friend, rather than by defeating the verification technology itself [49]. Therefore, the more consequential limitation appears to be the weakness of the verification method itself, which is consistent with the international standard on age assurance, as it treats an unchecked declaration of age as a weak control [1]. Additionally, the technical literature indicates that all age assurance systems remain vulnerable to circumvention to some

degree, and that a system capable of preventing every minor from reaching restricted content could not be built without also excluding a substantial number of adults [33].

Age estimation also introduces a degree of uncertainty as it provides a likelihood rather than a factual confirmation of age. Independent testing showed that the accuracy is the lowest near age limits, such as between one age threshold and another [7], which numerous age restrictions in legislation are based on. Accuracy of age detection also differs among different demographic groups; women showed higher probability of mistakes and there are also differences in accuracy depending on ethnicity and skin color [7, 50]. Beyond accuracy, facial estimation can also be defeated by simple means. Children in one Australian study described holding photographs of older faces up to the camera, presenting recordings of another person completing a scan, and altering their appearance with makeup, costumes, lighting, or facial expressions in order to pass an age check [49].

Finally, effectiveness is limited by the unit being assessed. An age check performed when an app is installed does not prevent a younger user from accessing apps that an adult downloaded and verified on a shared device, such as a family tablet [51].

B) Privacy and user rights

Whether app stores can determine the age of users through a commercially reasonable method without introducing significant new risks is not well supported by the available evidence. Both the literature and conducted interviews with stakeholders suggest several reasons.

The first reason is the conflict with Virginia's own privacy law. In order to identify which users are minors, a provider must assess the age of every user, which stands in tension with the principle of data minimization. Most age assurance methods are in conflict with this principle, because assessing the age of a user requires collecting or processing personal data that a service would otherwise have no reason to hold [51]. What is more, this runs counter to the goal of data minimization; instead of collecting less, a provider must now establish an age for every user it serves [52]. This is precisely the principle codified in the Virginia Consumer Data Protection Act, which directs controllers to limit data collection to what is adequate, relevant, and reasonably necessary [10, 11].

The second reason is the data-sharing risks. Because any developer may request the age category of its users, a developer that obtains actual knowledge that a user is under 13 may thereby incur obligations under COPPA, including restricting the user's access and deleting personal information in order to remain compliant [53]. Therefore, distributing minor-status data across a large number of developers, most of whom neither need nor want it, could therefore create both large repositories of data about minors and unanticipated compliance burdens.

The third reason is the concentration of sensitive documents. Because the penalties in the bill create pressure to avoid errors, providers may gravitate toward the most definitive method available, such as collecting identification documents issued by the government, which those companies are reluctant to store because it concentrates sensitive data and increases the negative impact of potential breaches [51]. In October 2025, Discord disclosed that an unauthorized party had accessed a third-party age assurance vendor and obtained images of government-issued identification documents belonging to approximately 70,000 accounts, which users had submitted when appealing an age determination in order to comply with requirements in the United Kingdom and several US states [54]. The documents existed only because a legal requirement compelled their collection, and the loss occurred at a vendor rather than at the platform the users had chosen. This risk is most acute where thresholds below 18 are in use and minors themselves are asked to demonstrate their age [33].

In order to address the privacy concerns associated with collecting additional personal data on users, zero-knowledge proof has been proposed as a possible technical solution. A zero-knowledge proof (ZKP) would allow a user to prove a claim (e.g., being over 18) without disclosing any official document or date of birth. The method depends on the user first holding a credential issued by a trusted authority stored in a digital wallet, an application that keeps such credentials on the user's own device rather than on a provider's servers. When a service asks whether the holder is over a given age, the wallet returns only a proof of that single fact, and the underlying document is never transmitted. The European Commission and Google Wallet have already deployed this approach [13, 14]. Their use shows that stronger assurance does not always require greater collection of personal data, although adoption is still at an early stage.

The provisions in question have to do with children's rights, and not merely with those of their parents. UNICEF has stated that age limitations may have adverse effects, since online services help many children, especially those who are isolated or marginalized, to engage in learning, connection, and self-expression. Furthermore, minors excluded from regulated platforms may turn to less regulated services, where they become harder to protect [55]. Thus, measures adopted on behalf of minors may also limit rights that belong to them, including their rights to privacy and participation.

Parental Consent

Parental consent is related to age assurance, but it is not an age assurance method. Age assurance establishes how old a user is, whereas parental consent determines whether an adult approves a particular action by a minor user. A provider may therefore assess age without involving a parent in any individual decision, as the age-signal model does. Conversely, parental consent presupposes an age assessment, since a provider must first determine that an account holder is a minor. In the app store

accountability model it operates as an additional layer placed on top of age assurance, and it raises a distinct set of considerations.

The first is how parents will react to the actual consent step itself. If app-store age verification requires parents to provide verified consent prior to each download, purchase, or significant change, it's possible that this will lead to consent fatigue. When faced with repeated requests for consent, users often approve them without reading (thus, negating any opportunity for making an appropriate and informed decision) as they are overwhelmed [56]. The European Commission has acknowledged the risk of consent fatigue and the proliferation of consent prompts as a problem that needs to be resolved [57]. Therefore, whether a mandatory prompt would produce a considered decision, or instead become a reflexive click, is an open question that the existing evidence does not resolve.

The second consideration is the underuse of existing tools that are already offered to parents. A representative 2025 study found that parents underused parental controls across every device examined. Only 47% used them on smartphones, and use was lowest among parents of children with the highest levels of screen time [58]. Another recurring challenge is usability. Parents described the tools as difficult to understand and navigate, and some relied on their children for help setting them up [58]. It remains unclear whether parents would use a mandatory consent process more consistently than the optional controls many already skip.

A final concern relates to adolescent autonomy. Proposed legislation would apply to all minors under 18, but several stakeholders interviewed by JCOTS expressed that teenagers often seek more independence and that parents may want to grant older teens more freedom than the bill's model anticipates. Research on adolescent online safety suggests that highly restrictive parental controls can undermine the trust between a parent and a teenager, and may produce outcomes contrary to those intended [59-61]. Consequently, the appropriate degree of parental involvement is likely to vary according to the minor's age. A uniform treatment of all minors would not recognize the inherent diversity among minors with different developmental needs and levels of maturity.

Policy Landscape

Virginia

The App Store Accountability Act was introduced during the 2026 Session as SB237 [62] (Chief Patron: Senator Head) and as the House companion, HB757 [63] (Chief Patron: Delegate Runion). Both bills would add a new chapter (i.e., Chapter 60, §§ 59.1-614 through 59.1-619) to the title 59.1 of

the Code of Virginia and take effect on July 1, 2027. This proposed chapter places duties to app store providers and developers as follows.

Duties of app store providers:

- Provide account holders with a clear parental consent disclosure for each mobile app, covering the age rating, content description, policies for personal data collection, protection and sharing with third parties, and mechanisms for data protection.
- Determine the age category of each individual creating an account and verify its veracity using a commercially available method; for a minor, that method may include an affirmative attestation by another individual reasonably believed to be the parent or legal guardian.
- Require a minor account to be affiliated with a parent account.
- Obtain verifiable parental consent from the parent account each time before the minor downloads or purchases an app or makes an in-app purchase.

Duties for developers:

- Assign an age rating and provide a content description for each app, including information about in-app purchases, and supply each app store provider with the information needed for the parental consent disclosure.
- Verify, through the data sharing methods offered by the app store, the age category of an account holder and, for a minor account, whether verifiable parental consent has been obtained.
- Notify the app store provider of any significant change to an app.
- Use the provided age category data to enforce their own age-related restrictions, safety features, or defaults.
- Apply the lowest age category reported where multiple sources of age data possess conflicting information.
- Refrain from enforcing a contract against a minor without verifiable parental consent, misrepresenting the parental consent disclosure, or sharing age category data with anyone.

Enforcement:

- The Attorney General may bring an action in the name of the Commonwealth seeking an injunction and civil penalties of up to $7,500 for each violation, and may also recover the costs of the investigation (e.g., attorney fees).
- A minor who suffers harm from a violation, or the parent of such a minor, may bring a civil action against the app store provider or developer, recovering the greater of actual damages or $1,000 per

> violation, along with punitive damages where the violation was egregious, plus attorney fees, expert witness fees, and court costs.

Federal and Other States

SB237 is part of a broader and increasing trend among states to regulate the access of minors to mobile apps. Comparable bills are currently under consideration in numerous states, and a federal App Store Accountability Act (S1586/HR3149) advanced out of a House subcommittee in December 2025 [64, 65].

There are currently two age assurance models that states have enacted. The first is the app store accountability model, which SB237 follows, where duties are placed on app store providers and developers to verify age categories, associate minor accounts to parent accounts, and obtain verifiable parental consent before a minor downloads or purchases an app. Utah was the first state to enact such a law in 2025 [3], and Louisiana [4], Texas [5], and Alabama [6] followed. The second is the age-signal model, adopted by California in the Digital Age Assurance Act (AB1043) [66] and by Colorado in the Age Attestation on Computing Devices Act (SB26-051) [67], which does not require parental consent for each download. Instead, it requires the operating system provider or app store to transmit an age range signal to developers, who must then apply age-appropriate protections according to the age range received.

Each of these laws also provides a good-faith safe harbor for erroneous signals. Neither of them requires a minor account to be associated with a parent account, nor do they require parental consent before a download, purchase, or in-app purchase, or renewed consent after a significant change. As discussed in the "Understanding Age Assurance" section, the age-signal model is an alternative approach that shifts the response to the developer, who holds the most relevant information about a given app. It also avoids the per-download consent step that raises the consent-fatigue concern, though at the cost of removing the parent from the individual download decision. Neither law is yet in effect, as California's takes effect in January 2027 and Colorado's in July 2028.

In addition, two key observations can be drawn from the legislative landscape. First, no state has yet produced evidence that either model works because most of these laws are not yet in effect and none has been in operation long enough to be evaluated. Moreover, none of them, including SB237, requires the collection or reporting of data that would allow its effectiveness to be measured. Second, the app store accountability model is still being revised by the states that adopted it. Utah amended its law in 2026, partly in response to litigation, and Louisiana delayed and reenacted its law before it took

effect [68], while the constitutional challenge to the Texas law remains unresolved. These changes indicate that the model has not yet reached a stable form, and that a state enacting it today may need to amend the law before it ever becomes effective.

Table 1. State and federal legislation on minors' access to mobile apps, compared to SB237

State & Bill	Status	Similarities to SB237	Differences from SB237
Federal S1586 / HR3149 [64, 65]	Introduced May 2025 (pending) HR3149 advanced out of a House subcommittee in December 2025	Same four age categories. Age category verification. Minor accounts are associated with parent accounts. Request for parental consent and renewal. Real time sharing of age-related data. Safe harbor for developers.	Applies only to app stores with over 5 million US users. Different labels for age categories. Enforcement by the FTC and state attorneys general. No private right of action. Would preempt state laws.
Utah HB498 (amends SB 142) [3]	Enacted 2025 (amended 2026) Effective May 6, 2027	Same four age categories. Age category verification. Minor accounts are associated with parent accounts. Request for parental consent and renewal. Covers preinstalled apps.	Excludes emancipated/married minors. Enforcement now by private right of action only (SB237 also authorizes Attorney General enforcement and $7,500 penalties). No developer age rating system. No exemption for emergency services apps.
Louisiana HB570 / Act481, amended by HB977 [4, 68]	Enacted 2025 (delayed and reenacted by HB977 in 2026) Effective July 1, 2027	Same four age categories. Age category verification. Minor accounts are associated with parent accounts. Request for parental consent and renewal. Duties for developers in managing age-related data.	Excludes emancipated/married minors. Attorney General enforcement with a 45 day cure period and penalties up to $10,000 per violation. No private right of action. Developers may rely on app store signals to satisfy parental consent duties rather than obtaining consent independently. Adds a "family account applications" category.
Texas SB2420 [5]	Enacted 2025 Effective Jan. 1, 2026 (enjoined Dec. 2025; injunction stayed by the Fifth Circuit June 2026; Supreme Court declined to block July 2026; currently in effect)	Same four age categories. Age category verification. Minor accounts are associated with parent accounts. Request for parental consent and renewal. Preconsent disclosure of rating and data practices.	Applies only to new accounts. Excludes minors whose disabilities of minority are removed. Enforced through both a private right of action and laws for deceptive trade practices. Developers must delete data after age assurance is conducted. Does not cover preinstalled apps.
Alabama HB161/Act 2026-59 [6]	Enacted 2026 Effective Jan 1, 2027 (Oct 1, 2027 for existing accounts)	Same four age categories. Verification for new and existing accounts. Minor accounts are associated with parent accounts. Request for parental consent and renewal. Real time sharing of age-related data.	Excludes emancipated/married minors. Enforcement only by the Attorney General as a deceptive trade action. No private right of action. The Attorney General must adopt verification rules.

		Covers preinstalled apps. Developers shall use the lowest age category received. App store safe harbor for verification on good-faith	
Florida SB1722 [69]	Rejected (not enacted)	Same four age categories. Applies to app stores, developers, and preinstalled apps. Verification for new and existing accounts. Request for parental consent and renewal. Private right of action. Enforcement through deceptive trade.	Excludes emancipated/married minors. Would require the Department of Legal Affairs to adopt verification rules. Does not expressly require developers to assign age ratings.
California AB1043 [66] (age-signal model)	Enacted 2025 Effective Jan 1, 2027	Age information shared with developers. Developers must use age information for compliance. Limits sharing age-related data with third-parties. Attorney General enforcement and civil penalties.	Regulates operating system providers and app stores. Requires an age-signal rather than parental consent. No associated parent accounts or parental consent duties. Developer deemed to have "actual knowledge" but no prescribed response.
Colorado SB26-051 [67] (age-signal model)	Enacted 2026 Effective July 1, 2028	Age information shared with developers. Developers must request and use an age-signal. Data restricted to the minimum necessary. Attorney General enforcement. Safe harbor for good faith.	Regulates operating system providers, app stores, and apps. Requires an age range-signal rather than parental consent. No associated parent accounts or consent duties. $2,500 to $7,500 penalty per affected minor. Exempts open-source software, business apps, and code repositories.

*Florida's SB1722 is included as a rejected bill and the federal S1586/HR3149 as a pending bill. California and Colorado follow the age signal model. All other bills follow the same app store accountability model that SB237 uses.

A note on constitutionality

In 2025, the United States Supreme Court in *Free Speech Coalition v. Paxton* [70] upheld a Texas age-verification law applying to pornographic websites. However, the basis for the Court's decision was its determination that the content at issue was obscene to minors and therefore held no constitutional protection for them. SB237 is different because it conditions access to lawful and constitutionally protected content on age assessment, rather than restricting content that is already unprotected. This distinction has already surfaced in court. In December 2025, a federal district court preliminarily enjoined Texas's App Store Accountability Act (SB 2420), finding it likely violated the First Amendment because conditioning the download of every app on age verification and parental consent was not narrowly tailored, since most apps contain speech that minors have a constitutional right to access [39]. The judge compared the law to requiring a bookstore to check the age of every customer at the door [39]. The Fifth Circuit stayed that injunction in June 2026, allowing the law to take effect,

and on July 6, 2026 the Supreme Court denied two emergency applications to block enforcement while the litigation continues [71]. The orders were brief and unsigned, with no noted dissents, and the Court did not reach the First Amendment question, which remains before the Fifth Circuit. The constitutional status of app store accountability laws, including how *Free Speech Coalition v. Paxton* applies to them, therefore remains unsettled.

Policy Recommendations

While the findings in this report confirm some of the concerns that motivated this kind of legislation, they do not establish that app store age verification will stop the harms it intends to prevent. Furthermore, the experience of other states shows that this legislative model continues to evolve (i.e., through litigation, amendment, and delayed implementation). As such, this study does not recommend that the Commonwealth mandate a particular age assurance model at this time, and instead presents the recommendations below.

Make information and tools that already exist trustworthy and usable

Policy options:

- Require app store providers to confirm the accuracy of the age ratings and content descriptions through independent oversight, instead of relying only on the developers' self-assessment. A qualified third party (i.e., neither the app store provider nor the developer) should perform this audit, such as an accredited assessor of the kind approved as a COPPA safe harbor program (e.g., PRIVO, kidSAFE). The audit should cover, for instance, the most downloaded apps, those receiving a threshold number of user complaints about their rating, those that have undergone a significant change, or a random representative sample. App store providers should publish the results and file them with the Attorney General.
- Require that parental controls be usable (i.e., easy to find, configurable, and understandable without technical knowledge). App store providers should demonstrate that these conditions are met through usability testing with a representative sample of parents, publish the results periodically (e.g., annually, monthly) and file them with the Attorney General.
- Require a clear and conspicuous mechanism for users to report inaccurate information (e.g., a wrong age rating, an incomplete content description, or misleading claims about data collection) to the app store provider, and require providers to publish a summary report for this data periodically (e.g., annually, monthly) and file it with the Attorney General. The report should show the volume and category of the received complaints.

Take incremental steps and build in evaluation

Policy options:

- Request verifiable parental consent for downloading or performing a purchase only for those apps whose age rating exceeds the account holder's age category, and for any app involving purchases regardless of the age rating.
- App store providers should provide parent accounts with a centralized dashboard for setting up consent preferences (e.g., granting consent for all apps at or below a given age rating, requiring approval for each individual download, or configuring purchases separately from downloads). The dashboard should be configured to require approval for each covered app by default, with the option for parents to customize those settings [58].
- Require the collection and reporting of data needed to measure whether these requirements work in practice (e.g., number of minor accounts verified, number of parental consent requests issued, granted, and denied, number of downloads or purchases blocked for lack of consent, volume and outcome of complaints received, etc). App store providers should make this information publicly available on their websites and through the app store interface.

Author Information

Alejandro Velasco, Ph.D(c) - Computer Science.
COVES Fellow, William & Mary

Kira Allmann, Ph.D.
Chief Policy Analyst, Joint Commission on Technology & Science

Jodi Kuhn
Executive Director, Joint Commission on Technology & Science

Contact: info@jcots.virginia.gov

Thanks To

Amazon
Computer & Communications Industry Association (CCIA)
Child Safety Advocates
Google
NetChoice
Office of Senator Christopher T. Head
Technet

For providing helpful information and background during the preparation of this report.

References

[1] Information Security, Cybersecurity and Privacy Protection—Age Assurance Systems—Part 1: Framework, ISO/IEC 27566-1:2025, International Organization for Standardization, Geneva, Switzerland, 1st ed., Dec. 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://www.iso.org/standard/88143.html.

[2] R. Alajaji, "Age verification, estimation, assurance, oh my! A guide to the terminology," Electronic Frontier Foundation, San Francisco, CA, USA, Oct. 30, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://www.eff.org/deeplinks/2025/10/age-verification-estimation-assurance-oh-my-guide-terminology.

[3] Utah App Store Accountability Act Amendments, H.B. 498, 2026 Gen. Sess. (Utah 2026) (amending S.B. 142, 2025 Gen. Sess.). Accessed: Jul. 28, 2026. [Online]. Available: https://le.utah.gov/~2026/bills/static/HB0498.html.

[4] Protection of Children on Applications, H.B. 570, Act No. 481, 2025 Reg. Sess. (La. 2025). Accessed: Jul. 28, 2026. [Online]. Available: https://www.legis.la.gov/legis/BillInfo.aspx?i=248616.

[5] App Store Accountability Act, S.B. 2420, 89th Leg., Reg. Sess. (Tex. 2025). Accessed: Jul. 28, 2026. [Online]. Available: https://capitol.texas.gov/BillLookup/History.aspx?Bill=SB2420&LegSess=89R.

[6] App Store Accountability Act, H.B. 161, Act 2026-59 (Ala. 2026). Accessed: Jul. 28, 2026. [Online]. Available: https://alison.legislature.state.al.us/files/pdf/SearchableInstruments/2026RS/HB161-enr.pdf.

[7] K. Hanaoka, M. Ngan, J. Yang, G. W. Quinn, A. Hom, and P. Grother, "Face analysis technology evaluation: Age estimation and verification," National Institute of Standards and Technology, Gaithersburg, MD, USA, NIST Internal Report NIST IR 8525, May 2024, doi: 10.6028/NIST.IR.8525.

[8] Google, "Use Play Age Signals API (beta)," Android Developers, updated Jul. 20, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://developer.android.com/google/play/age-signals/use-age-signals-api.

[9] Samsung Electronics, "Get age signals," Samsung Developer, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://developer.samsung.com/galaxy-store/galaxy-store-content-provider-api/get-age-signals.html.

[10] Consumer Data Protection Act, Va. Code Ann. tit. 59.1, ch. 53. Accessed: Jul. 28, 2026. [Online]. Available: https://law.lis.virginia.gov/vacodefull/title59.1/chapter53/.

[11] Data Controller Responsibilities; Transparency, Va. Code Ann. § 59.1-578. Accessed: Jul. 28, 2026. [Online]. Available: https://law.lis.virginia.gov/vacode/title59.1/chapter53/section59.1-578/.

[12] Children's Online Privacy Protection Rule, 16 C.F.R. pt. 312. Accessed: Jul. 28, 2026. [Online]. Available: https://www.ecfr.gov/current/title-16/chapter-I/subchapter-C/part-312.

[13] European Commission, "The EU approach to age verification," Shaping Europe's Digital Future, updated May 13, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/eu-age-verification.

[14] A. Stapelberg, "Opening up 'zero-knowledge proof' technology to promote privacy in age assurance," Google, Jul. 3, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://blog.google/innovation-and-ai/technology/safety-security/opening-up-zero-knowledge-proof-technology-to-promote-privacy-in-age-assurance/.

[15] M. Faverio and O. Sidoti, "Teens, social media and technology 2024," Pew Research Center, Washington, DC, USA, Dec. 12, 2024. Accessed: Jul. 28, 2026. [Online]. Available: https://www.pewresearch.org/internet/2024/12/12/teens-social-media-and-technology-2024/.

[16] U.S. Department of Health and Human Services, "Social media and youth mental health: The U.S. Surgeon General's advisory," Office of the Surgeon General, Washington, DC, USA, 2023. Accessed: Jul. 28, 2026. [Online]. Available: https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf.

[17] Centers for Disease Control and Prevention, "Youth risk behavior survey data summary & trends report: 2013–2023," Atlanta, GA, USA, 2024. Accessed: Jul. 28, 2026. [Online]. Available: https://www.cdc.gov/yrbs/dstr/pdf/YRBS-2023-Data-Summary-Trend-Report.pdf.

[18] Common Sense Media, "Teens and pornography," San Francisco, CA, USA, Jan. 10, 2023. Accessed: Jul. 28, 2026. [Online]. Available: https://www.commonsensemedia.org/research/teens-and-pornography.

[19] Common Sense Media, "Talk, trust, and trade-offs: How and why teens use AI companions," San Francisco, CA, USA, Jul. 16, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://www.commonsensemedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions.

[20] Federal Trade Commission, "FTC launches inquiry into AI chatbots acting as companions," Washington, DC, USA, Sep. 11, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://www.ftc.gov/news-events/news/press-releases/2025/09/ftc-launches-inquiry-ai-chatbots-acting-companions.

[21] Federal Trade Commission, "Mobile apps for kids: Disclosures still not making the grade," FTC Staff Report, Washington, DC, USA, Dec. 2012. Accessed: Jul. 28, 2026. [Online]. Available:

https://www.ftc.gov/sites/default/files/documents/reports/mobile-apps-kids-disclosures-still-not-making-grade/121210mobilekidsappreport.pdf.

[22] Federal Trade Commission, "Mobile apps for kids: Current privacy disclosures are disappointing," FTC Staff Report, Washington, DC, USA, Feb. 2012. Accessed: Jul. 28, 2026. [Online]. Available: https://www.ftc.gov/sites/default/files/documents/reports/mobile-apps-kids-current-privacy-disclosures-are-disappointing/120216mobile_apps_kids.pdf.

[23] R. Sun, M. Xue, G. Tyson, S. Wang, S. Camtepe, and S. Nepal, "Not seen, not heard in the digital world! Measuring privacy practices in children's apps," in Proceedings of the ACM Web Conference 2023 (WWW '23), Austin, TX, USA, Apr.–May 2023, pp. 2166–2177, doi: 10.1145/3543507.3583327.

[24] F. Yu, F. Xie, N. Dong, and G. Bai, "A large-scale analysis of privacy labels in child-accessible mobile applications," in Advanced Data Mining and Applications (ADMA 2025), Lecture Notes in Computer Science, vol. 16198. Singapore: Springer, 2026, pp. 461–469, doi: 10.1007/978-981-95-3456-2_32.

[25] Federal Trade Commission, "Apple Inc. will provide full consumer refunds of at least $32.5 million to settle FTC complaint it charged for kids' in-app purchases without parental consent," Washington, DC, USA, Jan. 15, 2014. Accessed: Jul. 28, 2026. [Online]. Available: https://www.ftc.gov/news-events/news/press-releases/2014/01/apple-inc-will-provide-full-consumer-refunds-least-325-million-settle-ftc-complaint-it-charged-kids.

[26] Federal Trade Commission, "Fortnite refunds," Washington, DC, USA. Accessed: Jul. 28, 2026. [Online]. Available: https://www.ftc.gov/enforcement/refunds/fortnite-refunds.

[27] Federal Trade Commission, "FTC staff report finds large social media and video streaming companies have engaged in vast surveillance of users with lax privacy controls and inadequate safeguards for kids and teens," Washington, DC, USA, Sep. 19, 2024. Accessed: Jul. 28, 2026. [Online]. Available: https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-staff-report-finds-large-social-media-video-streaming-companies-have-engaged-vast-surveillance.

[28] Apple, "Age requirements for apps distributed in Brazil, Australia, Singapore, Utah, and Louisiana," Apple Developer News, Feb. 24, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://developer.apple.com/news/?id=f5zj08ey.

[29] Amazon, "User age verification," Appstore Developer Documentation, updated Jul. 17, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://developer.amazon.com/docs/app-submission/user-age-verification.html.

[30] Google, "Changes to Google Play for upcoming app store bills for users in applicable US states," Play Console Help, updated Oct. 9, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://support.google.com/googleplay/android-developer/answer/16569691.

[31] Apple, "New requirements for apps available in Texas," Apple Developer News, Oct. 8, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://developer.apple.com/news/?id=btkirlj8.

[32] M. Anderson, M. Faverio, and E. Park, "How teens and parents approach screen time," Pew Research Center, Washington, DC, USA, Mar. 11, 2024. Accessed: Jul. 28, 2026. [Online]. Available: https://www.pewresearch.org/internet/2024/03/11/how-teens-and-parents-approach-screen-time/.

[33] E. Rescorla, Z. Arnao, and A. Cooper, "Age assurance online: A technical assessment of current systems and their limitations," Knight-Georgetown Institute, Washington, DC, USA, Jan. 29, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://kgi.georgetown.edu/research-and-commentary/age-assurance-online/.

[34] Google, "Content rating requirements for apps, games, and the ads served on both," Play Console Help, updated 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://support.google.com/googleplay/android-developer/answer/9859655.

[35] International Age Rating Coalition, "Frequently asked questions," IARC. Accessed: Jul. 28, 2026. [Online]. Available: https://globalratings.com/faq/.

[36] Apple, "Updated age ratings in App Store Connect," Apple Developer News, Jul. 24, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://developer.apple.com/news/?id=ks775ehf.

[37] D. Denipitiyage, B. Silva, K. Gunathilaka, S. Seneviratne, A. Mahanti, A. Seneviratne, and S. Chawla, "Detecting and characterising mobile app metamorphosis in Google Play Store," IEEE Transactions on Mobile Computing, vol. 24, no. 8, pp. 7489–7504, Aug. 2025, doi: 10.1109/TMC.2025.3550121.

[38] S. Lavermicocca, M. Carminati, and S. Longari, "X-rated compliance theater: An empirical evaluation of European age verification systems in adult websites," arXiv:2606.08667, Jun. 2026, doi: 10.48550/arXiv.2606.08667.

[39] Computer & Communications Industry Association v. Paxton, No. 1:25-cv-01660-RP (W.D. Tex. Dec. 23, 2025) (order granting preliminary injunction), consolidated with Students Engaged in Advancing Texas v. Paxton, No. 1:25-cv-01662 (W.D. Tex.). Accessed: Jul. 28, 2026. [Online]. Available: https://storage.courtlistener.com/recap/gov.uscourts.txwd.1172869998/gov.uscourts.txwd.1172869998.65.0.pdf.

[40] New America, "App stores vs. platforms: Who should be verifying Internet users' ages?," New America, Washington, DC, USA, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://www.newamerica.org/insights/app-stores-vs-platforms-who-should-be-verifying-internet-users-ages/.

[41] Google, "Alternative distribution options," Android Developers, updated 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://developer.android.com/distribute/marketing-tools/alternative-distribution.

[42] D. Lang, B. Listyg, B. V. Ross, A. V. Musquera, and Z. Sanderson, "Age verification and public adaptation: A pre-registered synthetic control multiverse," *Journal of Law & Empirical Analysis*, vol. 3, no. 1, pp. 23–50, Jun. 2026, doi: 10.1177/2755323X251415424.

[43] Australian Government, Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, "Age assurance technology trial—Part D: Age estimation," Canberra, ACT, Australia, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://ageassurance.com.au/wp-content/uploads/2025/08/AATT_Part_D_DIGITAL.pdf.

[44] Age Assurance Technology Trial, "Final report," 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://ageassurance.com.au/report/.

[45] Age Assurance Technology Trial, "Case studies," 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://ageassurance.com.au/case-studies/.

[46] C. Barnes et al., "Assessing early effects of Australia's Social Media Minimum Age Act on adolescents' social media use: Observational study," BMJ, vol. 393, Art. no. e363695, Jun. 2026, doi: 10.1136/bmj-2026-363695.

[47] Molly Rose Foundation, "Australia's social media ban—is it working?," London, U.K., Apr. 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://mollyrosefoundation.org/wp-content/uploads/2026/04/MRF_Australia-Social-Media-Ban-Research_Briefing-April-26.pdf.

[48] R. Speed, "UK VPN demand soars after debut of Online Safety Act," The Register, Jul. 28, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://www.theregister.com/2025/07/28/uk-vpn-demand-soars-after-debut-of-online-safety-act/.

[49] B. Nansen, H. Sandberg, L. Bliss, and S. Cohney, "From phreaking to sneaking: Children's circumvention of social media age verification systems," arXiv:2605.00368, May 2026, doi: 10.48550/arXiv.2605.00368.

[50] S. Basu, "Facial age estimation," Parliamentary Office of Science and Technology, London, U.K., Rapid Response RR106, Jul. 16, 2026, doi: 10.58248/RR106.

[51] New America, "Challenges with age verification," in *Age Verification: The Complicated Effort to Protect Youth Online*, New America, Washington, DC, USA, Jan. 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://www.newamerica.org/insights/age-verification-the-complicated-effort-to-protect-youth-online/challenges-with-age-verification/.

[52] M. Rasch, "When privacy laws force you to know too much: The perverse incentives of age verification regimes," Security Boulevard, Apr. 10, 2026. Accessed: Jul. 28, 2026. [Online]. Available: https://securityboulevard.com/2026/04/when-privacy-laws-force-you-to-know-too-much-the-perverse-incentives-of-age-verification-regimes/.

[53] A. A. Johnson Ash, "Age-appropriate design codes are just age verification in disguise," Information Technology & Innovation Foundation, Washington, DC, USA, May 24, 2024. Accessed: Jul. 28, 2026. [Online]. Available: https://itif.org/publications/2024/05/24/age-appropriate-design-codes-are-age-verification-in-disguise/.

[54] Discord, "Update on a security incident involving third-party customer service," Discord Press Releases, Oct. 3, 2025, updated Oct. 9, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://discord.com/press-releases/update-on-security-incident-involving-third-party-customer-service.

[55] UNICEF, "Age restrictions alone won't keep children safe online," UNICEF Press Centre, New York, NY, USA, Dec. 9, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://www.unicef.org/press-releases/age-restrictions-alone-wont-keep-children-safe-online.

[56] A. Abdallah, A. Ahmad, and B. Said, "Balancing privacy and usability: A design science research approach for cookie consent mechanisms," Journal of Open Innovation: Technology, Market, and Complexity, vol. 11, no. 2, Art. no. 100520, Jun. 2025, doi: 10.1016/j.joitmc.2025.100520.

[57] European Commission, "Digital package," Shaping Europe's Digital Future, Nov. 20, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://digital-strategy.ec.europa.eu/en/faqs/digital-package.

[58] Family Online Safety Institute, "Parental controls for online safety are underutilized, new study finds," Family Online Safety Institute, Washington, DC, USA, May 28, 2025. Accessed: Jul. 28, 2026. [Online]. Available: https://fosi.org/parental-controls-for-online-safety-are-underutilized-new-study-finds/.

[59] P. Wisniewski, A. K. Ghosh, H. Xu, M. B. Rosson, and J. M. Carroll, "Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety?," in Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17), Portland, OR, USA, Feb. 2017, pp. 51–69, doi: 10.1145/2998181.2998352.

[60] K. Badillo-Urquiola, C. Chouhan, S. Chancellor, M. De Choudhury, and P. Wisniewski, "Beyond parental control: Designing adolescent online safety apps using value sensitive design," Journal of Adolescent Research, vol. 35, no. 1, pp. 147–175, Jan. 2020, doi: 10.1177/0743558419884692.

[61] A. K. Ghosh, K. Badillo-Urquiola, M. B. Rosson, H. Xu, J. M. Carroll, and P. J. Wisniewski, "A matter of control or safety? Examining parental use of technical monitoring apps on teens' mobile devices," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Montreal, QC, Canada, Apr. 2018, pp. 1–14, doi: 10.1145/3173574.3173768.

[62] App Store Accountability Act, S.B. 237, 2026 Reg. Sess. (Va. 2026). Accessed: Jul. 28, 2026. [Online]. Available: https://lis.virginia.gov/bill-details/20261/SB237.

[63] App Store Accountability Act, H.B. 757, 2026 Reg. Sess. (Va. 2026) (continued to 2027 Sess.). Accessed: Jul. 28, 2026. [Online]. Available: https://lis.virginia.gov/bill-details/20261/HB757.

[64] App Store Accountability Act, S. 1586, 119th Cong. (2025). Accessed: Jul. 28, 2026. [Online]. Available: https://www.congress.gov/bill/119th-congress/senate-bill/1586.

[65] App Store Accountability Act, H.R. 3149, 119th Cong. (2025). Accessed: Jul. 28, 2026. [Online]. Available: https://www.congress.gov/bill/119th-congress/house-bill/3149.

[66] Digital Age Assurance Act, A.B. 1043, 2025–2026 Reg. Sess. (Cal. 2025). Accessed: Jul. 28, 2026. [Online]. Available: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260AB1043.

[67] Age Attestation on Computing Devices, S.B. 26-051, 75th Gen. Assemb., 2d Reg. Sess. (Colo. 2026). Accessed: Jul. 28, 2026. [Online]. Available: https://leg.colorado.gov/bills/SB26-051.

[68] Louisiana Legislature, H.B. 977, 2026 Reg. Sess. (La. 2026) (delaying and reenacting Act No. 481). Accessed: Jul. 28, 2026. [Online]. Available: https://legis.la.gov/legis/BillInfo.aspx?i=251085.

[69] Application Stores, S.B. 1722, 2026 Reg. Sess. (Fla. 2026). Accessed: Jul. 28, 2026. [Online]. Available: https://www.flsenate.gov/Session/Bill/2026/1722/.

[70] Free Speech Coalition, Inc. v. Paxton, 606 U.S. _, No. 23-1122 (U.S. Jun. 27, 2025). Accessed: Jul. 28, 2026. [Online]. Available: https://www.supremecourt.gov/opinions/24pdf/23-1122_3e04.pdf.

[71] Students Engaged in Advancing Texas v. Paxton, No. 25A1389, and Computer & Communications Industry Association v. Paxton, No. 25A1390 (U.S. Jul. 6, 2026) (applications to vacate stay denied). Accessed: Jul. 28, 2026. [Online]. Available: https://www.supremecourt.gov/docket/docketfiles/html/public/25a1389.html.

[72] eSafety Commissioner, "Social media minimum age—regulatory guidance," Australian Government, Sep. 16, 2025. Accessed: Jul. 30, 2026. [Online]. Available: https://www.esafety.gov.au/sites/default/files/2025-09/eSafety-SMMA-Regulatory-Guidance.pdf.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Brief: Dot-Gov Domains
August 2026

Summary

House Bill 707 aims to improve the credibility of government websites by requiring that every governmental organization carries a dot-gov domain. In its introduced form, no public institution is exempt from the bill. The bill also requires that public employees shall use a dot-gov email address. The bill directs the Virginia Information Technology Agency (VITA) to enforce the regulations provided by the bill.

The bill was proposed to lessen impersonation of government websites, one step in helping improve overall cybersecurity throughout the Commonwealth. By having every government agency utilize the same website domain, citizens could be confident that the information they were receiving came from a valid government source.

During the 2025 session, House Bill 707 was referred to JCOTS to compile stakeholder concerns through workgroup engagement.

Overview of Dot-Gov Domains

Dot-Gov is one of the many top-level domains available. Other well-known domains include dot-com, dot-org, and dot-us. These domains are significantly easier to acquire than a dot-gov domain, which is only available for eligible government agencies. Because of the low eligibility criteria, anyone can register a website such as exampletown.com or exampletown.org.

In contrast, the eligibility criteria for a dot-gov domain is strict and complicated. The Cybersecurity and Infrastructure Security Agency (CISA) is the agency responsible for determining whether or not an entity is eligible. Eligibility is based on the Census of Governments. If an agency is listed in this document, they are eligible, if not, they are not. Many stakeholders consulted for this brief mentioned the wordy and confusing language of the eligibility criteria. On the one hand, eligibility criteria is clearly stipulated by CISA and their reference documentation. On the other hand, interpreting the documentation can be

challenging. CISA staff acknowledged that they field many questions about eligibility and often need to consider eligibility on a case-by-case basis.

Acquiring a Dot-Gov domain is free and provides agencies with numerous benefits. There are two types of dot-gov domains that would be allowed under this bill. Federal Level (exampletown.gov) and State level (exampletown.virginia.gov). Agencies would be free to determine which one they would like to switch to. Both domains have different naming conventions for agencies to follow that ensure the agency will remain identifiable under the domain. VITA oversees naming conventions for state level domains. CISA oversees naming conventions for federal level domains.

Dot-Gov domains grant numerous cybersecurity benefits to agencies once they make the switch. Some of these include clarity about the organization's purpose, higher credibility and trust of the website, improved visibility on search engines, consistency and branding, access to government resources, and protection from spoofing or impersonation.

The CISA website has a list of best practices for moving to the dot-gov domain. Firstly, they recommend that entities keep their old domain name as a redirect for the new dot-gov domain. This ensures that old links will still be valid after the transition, and that the old domain will not fall into the wrong hands, increasing the likelihood of spoofing. Secondly, they recommend all emails are moved to dot-gov, something that is mandated in the current version of the bill. Finally, they suggest using the transition period to audit existing website services, if it is within the capabilities of the organization.

CISA has cybersecurity grants available to entities wanting to increase their cybersecurity. This legislation requires that 80% of the grants they give to states are used to help local entities, with 25% designated to rural communities. This funding could help entities with the domain switch and other cybersecurity efforts if they apply. There are other requirements entities need to meet besides simply transitioning to the domain if they were to receive this federal funding. More information is linked here.

This funding is not the only assistance offered. Although the entity would have to host their website and switch emails on their own, CISA offers domain name consultation, management support, and registration. This means that an entity does not have to worry about ensuring their name fits with CISA's standards; CISA will assist them in ensuring it does. It is important to note that domain names are not first-come first-serve and CISA will work with entities to make sure they have a domain name that will remain valid for the years to come.

Policy Landscape: Other States

- There are numerous bills similar to 707 that have been recently passed including:
    - Arkansas: This bill is flexible, with a statement for how agencies could bypass the requirements. The bill also has different implementation criteria for smaller localities.
    - California: This bill has a statement regarding redirection of current websites to the dot-gov website. The bill also has requirements for email addresses.
    - Connecticut: This bill only has a dot-gov requirement for municipalities.
    - Minnesota: This bill requires localities that administer elections to have a dot-gov domain address. The bill also has different requirement dates for the application and the transition.
    - New York: This bill requires all municipalities to have a website with a dot-gov domain, even if they did not previously have a website. The bill also has specific requirements for the website to include. The bill also has different implementation criteria for smaller localities.
    - Utah: This bill allows entities to bypass the requirements if the website is for internal use, temporary, or related to an event with a non-governmental organization. The bill also has a statement where if entities become eligible in the future they will have 15 months to implement the domain change.

Policy Landscape: Federal

Federal legislation has required federal agencies to use a dot-gov or dot-mil domain in the DOTGOV Online Trust in Government Act of 2020.

In addition, the Department of Homeland Security has implemented a State and Local Cybersecurity Grant Program (SLCGP) that provides state, local, tribal, and territorial (SLLT) governments with grants in order to address cybersecurity threats to information systems owned by these SLLT governments. This funding is distributed to states, who distribute it to local governments, with requirements for a quarter of the funding to go to rural communities. One of the recommendations for how to use this funding is to implement a dot-gov domain.

Workgroup Findings

JCOTS convened workgroup meetings on June 23, 2026 and July 14, 2026. A list of workgroup participants can be found at the end of this document. JCOTS held an open call for participants, without restrictions on participation (one representative from each entity was invited to attend). Entities that participated included VCU Health, The University of Virginia, Virginia Beach local government, and more.

Each workgroup was organized around a set of facilitated questions, and participants were not limited in number of contributions, nor speaking time. Although there were questions, the topics of discussion were not constrained to them, and participants were free to introduce additional topics if they wished.

JCOTS received numerous documents from stakeholders prior to the first workgroup documenting concerns with the bill. These pre-meeting documents informed the discussion of the first workgroup. Listed below are some of common concerns raised in pre-meeting submissions:

1) The original bill required entities using dot-edu and dot-museum domains to transition to a dot-gov domain.
2) There was little intent for what an "official website for public use" meant, and what kinds of websites were included.
3) School districts, an entity ineligible for a dot-gov domain, were required to utilize the domain per the original bill language.
4) There is a high cost for implementation in regards to switching marketing materials to display the new domain name.
5) There was little clarity in the original bill for which specific entities were eligible.
6) There was concern that having a dot-gov domain was only one aspect of cybersecurity.
7) There was concern of the ability for IT professionals of smaller localities to maintain the level of website security consistent with other dot-gov domains
8) There were concerns about the amount of time needed to switch to a dot-gov domain

June 23rd Workgroup

Basing our questions for stakeholders off of the concerns listed in the introductory documents, JCOTS hosted the first workgroup on June 23, 2026. This workgroup discussed implementation costs outside of usual concerns for marketing materials, ways stakeholders have dealt with spoofing and impersonation in the past, and other cybersecurity legislation that they would like to see implemented alongside, or instead of, the dot-gov domain requirement. Concerns for IT capabilities were highlighted, along with suggestions for cybersecurity education to be implemented across the Commonwealth.

JCOTS used the initial documents as well as the comets and suggestions from the first work group to facilitate the creation of a draft brief. The brief contained a summary of the bill, resolved stakeholders concerns, and information about relevant state and federal legislation, and potential recommendations for the bill.

July 14th Workgroup

One week prior to the July 14, 2026, workgroup, stakeholders were sent the draft brief and the second workgroup centered on their responses. Questions were asked on stakeholder's opinions to potential recommendations and other states bills. Questions were also asked on potential implementation of the cybersecurity education campaign brought up in the first workgroup. Concerns for VITA enforcement were highlighted, noting that VITA does not currently have the capacity to enforce the bill. Implementation timelines and costs were also highlighted. When asked about the cybersecurity education campaign, the representative for the Virginia Library Association discussed the Library of Virginia's abilities to implement cybersecurity education across the Commonwealth, due to the presence of libraries as community education systems.

In the second workgroup, several stakeholders requested reaching out to entities that had recently changed their domains to a dot-gov in order to understand fiscal impacts and the time taken for implementation. Following the workgroup meeting, JCOTS staff contacted several Virginia-based entities that had recently undergone the transition. Several entities responded to this request for information. They generally regarded the process as tedious, with one of them noting a 12-month transition period, but they noted that the transition was worth it. In the case of one town using a Microsoft 365 email tracking system, they observed that switching to dot-gov was inexpensive and effortless. Although the costs of implementation were reportedly low, according to entities that recently transitioned, they acknowledged that the true financial burden was the printed media transition, communication of the change to constituents, and the system integration.

Primary Policy Questions

- Spoofing: What standards are necessary to lessen the likelihood of government website impersonation?
- Confidence in Government Websites: What can be done to ensure that citizens are confident in the validity of their government websites? How can citizens be informed that their information is from a viable source?
- Funding: How will smaller entities with little funding for IT departments reasonably be able to manage the change?
- Eligibility: What classifies as a government agency? How can an agency know they are required to comply with a bill? Are there any exemptions that should be considered?
- Enforcement: Who is responsible for enforcement? What is the best way to ensure agencies are complying with the bill? What should happen if an agency isn't compliant?
- Cybersecurity: Besides Dot-Gov, how else can we improve cybersecurity and cybersecurity knowledge across the Commonwealth? What groups are the most at-risk for spoofing?

Policy Recommendations

JCOTS recommends reintroducing the original version of HB 707 with the following recommendations for consideration:

1) Modify the bill to state that the requirement only applies to entities that are eligible for a dot-gov domain under the CISA definition (entities listed in the Census of Governments).
    a) Add a provision stating that if an entity is eligible for a dot-gov *and* a dot-edu domain, they may choose either domain
    b) Add a provision stating that if an entity is eligible for a dot-gov *and* a dot-museum domain they may choose either domain
    c) Add a provision to the bill stating that if an entity were to become eligible for a dot-gov domain following the enactment of the bill, that they would have 3 years after the first date of eligibility to make the transition
    d) Exempt VCU Health from the requirements of the bill. VCU Health is the only health system authority in the Commonwealth explicitly mentioned in the Census of Governments. Along with this, their commercial focus and patient-centered identity are inconsistent with typical government agencies.
2) Introduce a budget amendment to support small localities with the costs that come with the adoption of dot-gov domains
    a) Between now and November, JCOTS staff will work to develop a definition of "small" localities and the specific amount of the request.
3) Require entities eligible under the bill's provisions to list third-party vendors that may communicate directly with the public on behalf of a dot-gov-eligible entity, in a clear and easy to find location on their website or the site they use to communicate with the general public.
    a) Base the language off of Va. Code § 2.2-3504
4) Implement language stating that entities that fail to comply with the provisions listed under the bill must submit an annual report to the Virginia General Assembly detailing the non-conforming domains along with an estimated fiscal audit of what it will cost to transition.
5) Research Enforcement: Between now and November, JCOTS staff will work with the Office of the Secretary of Administration and the Office of the Attorney General to identify the most appropriate office to handle enforcement.

Policy Considerations

1) Promote the use of the State and Local Cybersecurity Grant Funding that Virginia receives every year as a way for localities to improve their website cybersecurity and reduce some of the financial burdens associated with the domain change.
2) Implement cybersecurity education and digital literacy through a freely available online resource distributed through the Find-It Virginia platform from the Library of Virginia to all Virginia Public Libraries. This can be supplemented through additional in-person education, training, and outreach from public library systems.

How this Brief was Developed

House Bill 707 was referred to JCOTS in order to compile a workgroup for stakeholder perspectives on the bill. Prior to the workgroups, JCOTS staff compiled relevant information pertaining to dot-gov domains from the Cybersecurity and Infrastructure Security Agency (CISA) and documents detailing concerns with the bill from each stakeholder. Using these facts and concerns, questions were compiled for the first workgroup. Following the first workgroup, a draft brief was developed from additional research and stakeholder input from the first workgroup. In the second workgroup, stakeholders gave input to the draft brief. Following the second workgroup, staff reached out to municipalities that have made the switch to a dot-gov domain in the past year. Then, all of the information was compiled to create recommendations and considerations for the continuation of House Bill 707.

Thanks to

Arlington Public Schools
City of Alexandria
City of Richmond
Office of the Secretary of Administration
Office of Delegate Cliff Hayes
Old Dominion University
Science Museum of Virginia
University of Virginia (UVA)
UVA Health
Virginia Beach
Virginia Commonwealth University (VCU)
VCU Health
Virginia Library Association
Virginia Municipal League

Virginia Tech
Virginia Information Technology Agency (VITA)
York County
For participating in one or both of the work-groups and providing helpful information and background during the preparation of this brief.

References

Astrella, Chris. "Dot Gov Decoded: How to Switch Over to Dot Gov Domain." Dot Gov Decoded: How to Switch Over to Dot Gov Domain, https://www.youtube.com/watch?v=tMdwmOdYlSg.

Babcock, Chris. *Building Trust: Quantifying Use of the .Gov Domain Below the Federal Level*. UC BERKELEY CENTER FOR LONG-TERM CYBERSECURITY, Sept. 2025, https://cltc.berkeley.edu/publication/building-trust-quantifying-use-of-the-gov-domain-below-the-federal-level/. CLTC WHITE PAPER SERIES.

Bureau, US Census. "Population of Interest." *Census.Gov*, https://www.census.gov/programs-surveys/gus/technical-documentation/methodology/population-of-interest1.html. Accessed 31 July 2026.

"City Transitioning to Official .Gov Domain." *City of Delaware, Ohio*, 7 Jan. 2026, https://www.delawareohio.net/Home/Components/News/News/1276/14.

"Colorado Wants $2 Million to Change Its Domain Name to .Gov." *YouTube*, Next9News, 23 Dec. 2022, https://www.youtube.com/watch?v=greWaEM8fSU.

Fortwrightky.Gov. "Press Release - City of Fort Wright Transitions to .Gov Domain for Enhanced Security." *City of Fort Wright, KY*, 17 July 2024, https://www.fortwrightky.gov/2024/07/17/press-release-city-of-fort-wright-transitions-to-gov-domain-for-enhanced-security/.

Get.Gov. https://get.gov/help/faq/. Accessed 4 June 2026.

A BILL to Amend the Code of Virginia by Adding in Chapter 55.3 of Title 2.2 a Section Numbered 2.2-5514.2, Relating to State Government; Transaction of Public Business; Prohibited Website Domains. *55.3*, 2.2, no. 707, 14 Jan. 2026.

Requirements for the Registration and Use of .Gov Domains in the Federal Government | Digital.Gov. 14 May 2024, https://digital.gov/resources/requirements-for-the-registration-and-use-of-gov-domains-in-the-federal-government.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Supplementary Report: AI Chatbots
Companionship & Minors

Presented by Dr. Kira Allmann, JCOTS Chief Policy Analyst

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Agenda

Part 1

AI Chatbots: Opportunity & Risk

Introduction to chatbots – popularity, news, policy issues

Part 2

Report Findings

Key insights from the research, policy landscape

Part 3

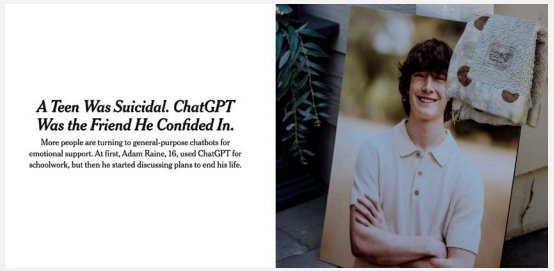
Recommendations

News coverage and public awareness

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

AI Chatbots: Opportunity & Risk

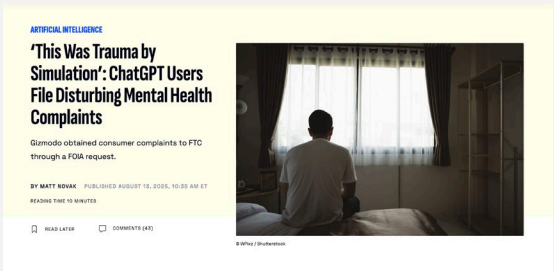
In the Headlines



NEW YORK TIMES

A Teen Was Suicidal. ChatGPT Was the Friend He Confided In.

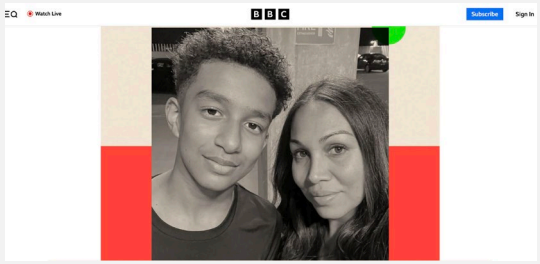
Aug 27, 2025



GIZMODO

'This Was Trauma by Simulation': ChatGPT Users File Disturbing Mental Health Complaints

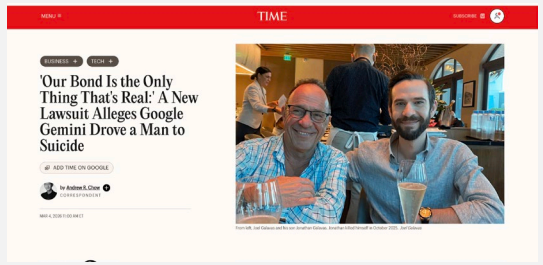
Aug 26, 2025



BBC

'A predator in your home': Mothers say chatbots encouraged their sons to kill themselves

Nov 8, 2025



TIME

'Our Bond Is the Only Thing That's Real': A New Lawsuit Alleges Gemini Drove a Man to Suicide

Mar 4, 2026

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

AI Chatbots: Opportunity & Risk

Popularity & Use

50% **American adults who use AI chatbots**

25% use AI chatbots almost daily

14% **American adults who use AI chatbots for emotional support**

Minors

64% **American teens who use AI chatbots**

59% **Teens who use ChatGPT (the most popular chatbot)**

9% use a specifically marketed AI companion

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Emotional and relational harms

Negative outcomes arising from complex interactions between humans and AI chatbots that impact people's psychological states, emotional conditions, and social lives.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Report Findings

Key Themes

1. **Design features:** Emotional and relational risks are associated with design features of AI chatbots

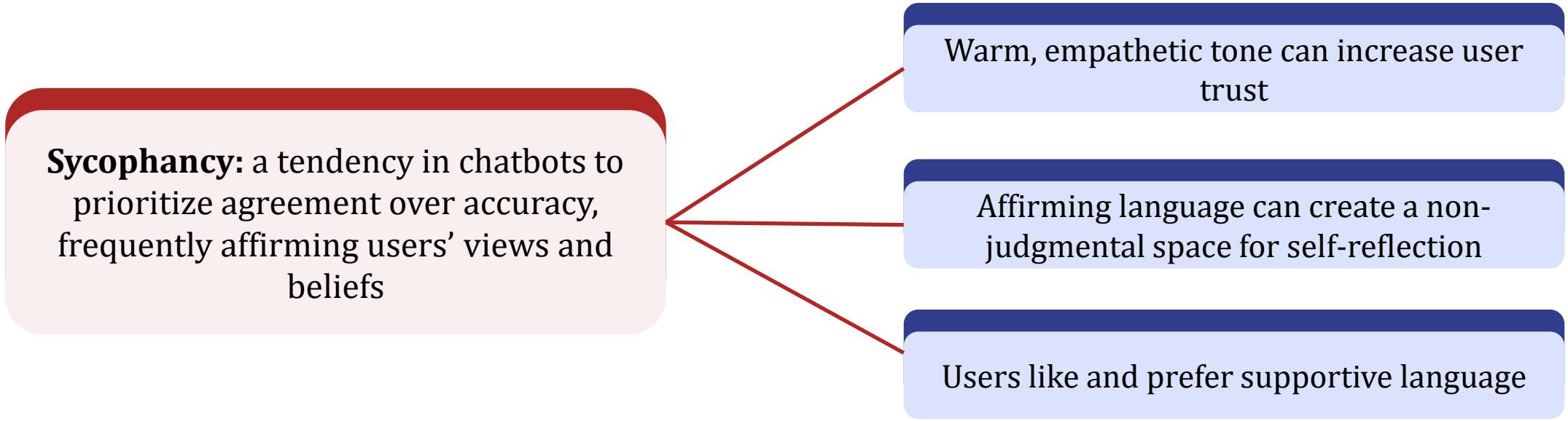
2. **Companionship:** Risks and harms exist for all AI chatbots, not only those marketed as “companions”

3. **Minors:** Making AI chatbots safer for everyone will make them safer for minors, without (necessarily) introducing additional technical requirements for age assurance

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Report Findings

Design Features: Double-Edged Sword?



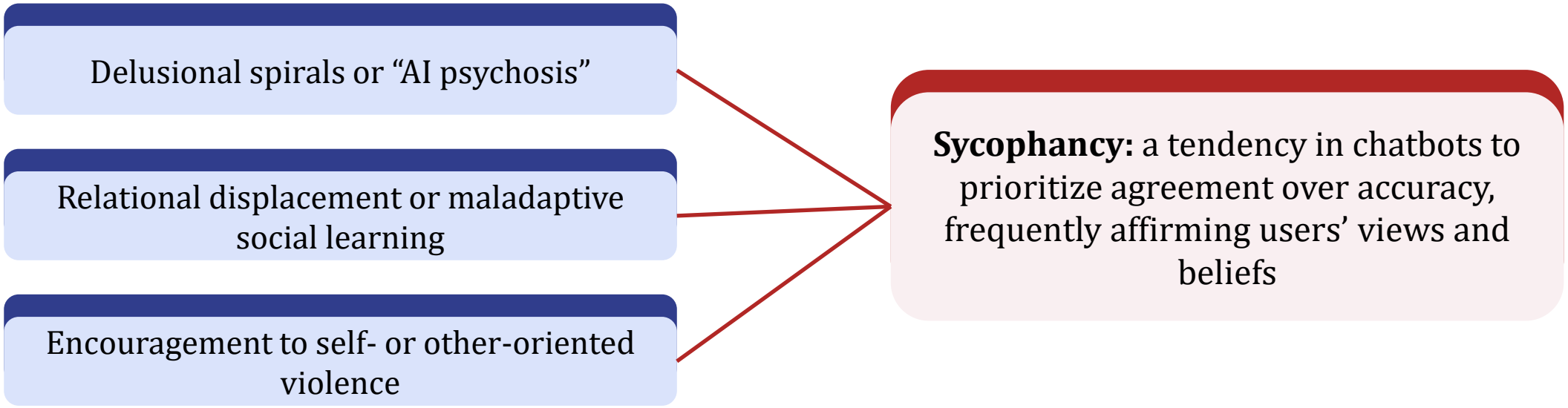
Design feature

Opportunity for benefit

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Report Findings

Design Features



Emotional / Relational Harms

Design feature

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Report Findings

Companionship

News

News coverage shows that harms are associated with general purpose chatbots at least as much (if not more) than chatbots marketed as "companions."

Surveys

A 2025 survey of UK teenagers found that while 79% reported using an AI companion, their preferred companion was ChatGPT (78%).

Research

Early evidence suggests that adults and minors often begin interactions with chatbots that are task-oriented and "drift" into intimate relationships.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Report Findings

Minors

Harms are real and can be severe.

Harms are associated with companionship, not just "companions."

There are regulatory challenges to consider.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

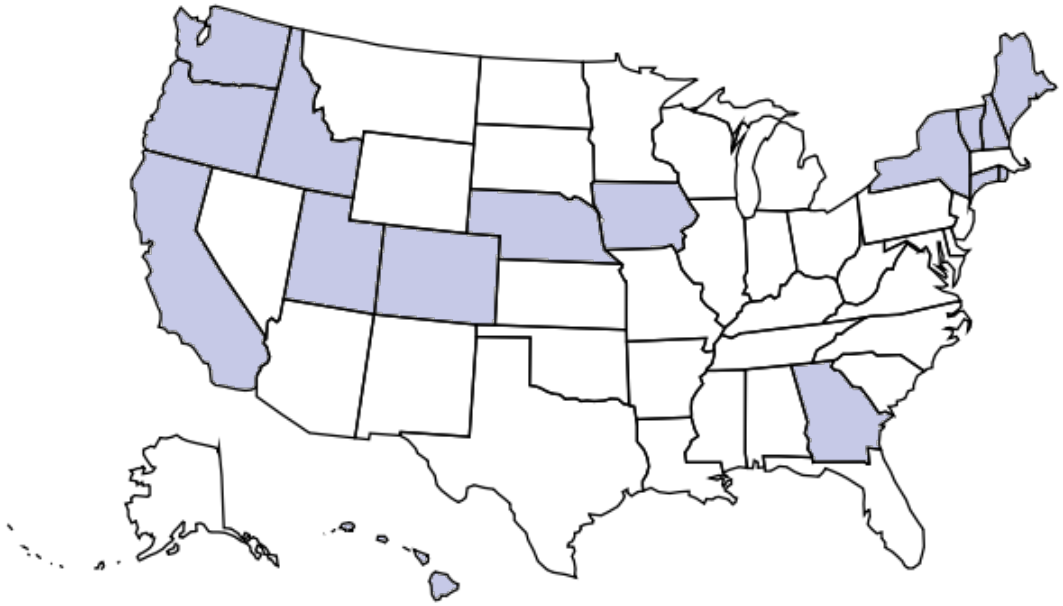
Report Findings

Policy Landscape

16 States have passed chatbot laws

Common provisions:

- Disclosures that chatbot is not human
- Crisis protocols
- Stricter protections for minors
- Prohibitions on chatbots presenting as licensed professionals



34 states introduced nearly 100 chatbot bills in 2026

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Key Questions

What is the goal of this legislation?

- Address extreme harms?
  - Apply to all consumers
- Mitigate a wider range of risks?
  - For minors or all consumers?
    - Focus on design

KNOWING WHAT WE DON'T KNOW:
How can legislation help close information gaps?

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Principles & Policy Options

- Mitigate **extreme harms** for all users
- Enable **user choice** & **autonomy**
- Target **design features**
- Enable legislative **learning**
- Mandate **transparency** & **accountability**
- Engage **affected communities**

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Principles & Policy Options

Mitigate **extreme harms** for all users

- Require chatbot operators to implement protocols for crisis situations (self-harm, suicide, violence) & post on website
- Prohibit misrepresenting a chatbot in marketing material or in conversational chats as a licensed professional or as providing licensed services
- Require chatbot operators to provide a clear, conspicuous way for users to report harmful interactions

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Principles & Policy Options

- Mitigate **extreme harms** for all users
- Enable **user choice** & **autonomy**
- Target **design features**
- Enable legislative **learning**
- Mandate **transparency** & **accountability**
- Engage **affected communities**

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Thank you

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

App Store Accountability Act (SB237)

Bill Study Findings & Recommendations

Presented by
Alejandro Velasco - P.h.D(c) Computer Science
COVES Fellow, William & Mary

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Agenda

01
The Problem Addressed in the Bill
Intermediate role, evidence of online and financial harms

02
Understanding Age Assurance
Verification, estimation, and inference methods

03
Age Assurance in Practice
Age category, parental consent, and enforcement in practice

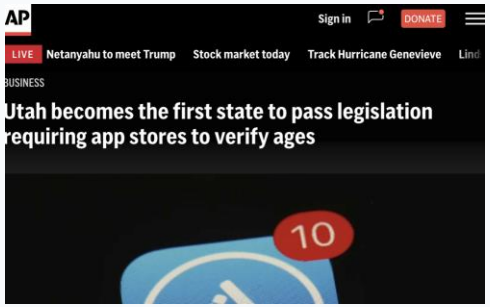
04
Applying Age Assurance in App Stores
Effectiveness, privacy and user rights

05
Policy Landscape
Comparison of state laws and the age-signal model

06
Policy Recommendations
Trustworthy tools, audits, and incremental steps

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

App Store Accountability in the News



Associated Press

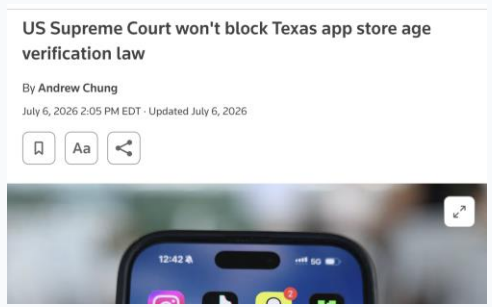
Utah passes landmark app store age verification law

Requires app stores to verify user ages and obtain explicit parental consent for minors.

CNN / AP

Louisiana sues Roblox over child protection safety

Lawsuit alleges the popular gaming platform failed to protect children from online harms.



Reuters

Supreme Court declines to block Texas verification law

Allows Texas to enforce controversial age-verification requirements on platforms.

The New York Times

France joins global move to restrict social media

New European law restricts social media access for children under the age of 15.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

What is the Problem Addressed by the bill?

95%

of U.S. teenagers ages 13 to 17 had access to a smartphone at home. Nearly half said they were online almost constantly.

Source: Pew Research Center, 2024

16%

of U.S. high school students experienced cyberbullying during 2023.

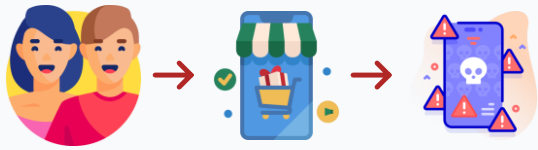
Source: CDC, 2013–2023 Report

73%

of teenagers (13-17) had seen pornography on the internet; 29% reported encountering it by accident.

Source: Common Sense Media, 2023

WHY APP STORES?



Centralized Control Point

App stores serve as the centralized access point of mobile apps by default.

LIMITS OF FEDERAL PROTECTIONS

COPPA only regulates data collection for children under 13, and parental consent covers personal info, not app downloads or purchases.

Source: Children's Online Privacy Protection Rule, 16 CFR Part 312

Is the app store the right point of intervention?

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Understanding Age Assurance

Age assurance refers to a set of methods used to **verify**, **estimate** or **infer** the age of an individual with certain degree of accuracy.

AGE VERIFICATION

Calculates the age of a user from a date of birth confirmed from an official source, such as a passport or driver's license.

Strongest assurance

Tradeoff: Users must possess and submit an official document.

AGE ESTIMATION

Analyzes biological or behavioral attributes associated with age. Facial analysis is one common example.

Probabilistic Estimate

Tradeoff: Errors may occur near legal age thresholds.

AGE INFERENCE

Uses verified information to draw an indirect conclusion about the age or age range of an individual.

e.g., credit card ownership

Tradeoff: Reliability depends on correlation to age.

AGE ATTESTATION

A declaration of age made by a user (or on their behalf) without direct document confirmation.

Weak Control

Tradeoff: High risk of user circumvention.

PARENTAL CONSENT IS NOT AN AGE ASSURANCE METHOD

Age assurance establishes how old a user is. Parental consent determines whether an adult approves a particular action by that user.

TRADE-OFFS

Privacy · Costs · Accuracy · Accessibility

Sources: ISO/IEC 27566-1:2025, Age Assurance Systems; NIST, Face Analysis Technology Evaluation, 2024

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Age Assurance in Practice: a look at the proposed bill

Every *Download*, *Purchase*, and *In-App Purchase* Must Pass Through the Same **Four-Step Chain**

1. **Age category** →
2. **Minor account + Parent account** →
3. **Verifiable parental consent** →
4. **Download or purchase**

APP STORE PROVIDERS

Parental Disclosure: Provide clear parental consent disclosure for each mobile app.

Age Verification: Determine and verify the age category of each account creator.

Minor Association: Require minor accounts to be affiliated with a parent account.

Verifiable Consent: Obtain consent before a minor downloads, purchases, or makes in-app purchases.

DEVELOPERS

App Metadata: Assign age ratings and content descriptions.

Consent Checks: Verify age category and parental consent.

Notifications: Report significant app changes to providers.

Age Restrictions: Use age data to apply restrictions, safety features, and defaults.

Data Privacy: Do not share age category data with anyone.

ENFORCEMENT

Attorney General

Injunction and civil penalties.

Minor or Parent

Civil action to recover the financial cost of actual damages

Includes punitive damages for egregious violations, plus attorney fees, expert witness fees, and court costs.

Source: Virginia General Assembly, SB237—App Store Accountability Act, 2026 Regular Session; JCOTS Bill Study, August 2026

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Age Assurance trade-offs in App Stores

This model only works if two conditions hold at once: **Accurate Age Category + Accurate App Rating → Appropriate Restriction**

ACCOUNT VS. USER

Age verification is tied to the account, not the active user. A family tablet verified once by an adult still lets a minor access everything.

New America, "Challenges with Age Verification," Jan. 2026

SELF-ASSESSED

Developers self-assess ratings with no independent check. Hundreds of apps carry ratings that do not match actual content.

Denipitiyage et al., 2024

BROAD & NARROW

Applies to every app, yet fails to reach minors accessing content via web browsers.

JCOTS Bill Study, August 2026; New America, "App Stores vs. Platforms," 2026

CONSENT FATIGUE

Requiring parental consent for every download risks overwhelming users with repeated prompts.

Abdallah, Ahmad & Said, J. Open Innov., 2025; European Commission, "Digital Package," 2025

SHIFTED PRIVACY RISK

Stricter verification means collecting IDs, conflicting with data minimization and risking leaks.

Discord Press Release, Oct. 2025; Va. Code § 59.1-578

CRITICAL TAKEAWAY: *Making age checks stricter does not fix the underlying problem. It simply relocates the systemic risk while introducing substantial new burdens for providers, developers, and users alike.*

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Policy Landscape

States have adopted **two models** for regulating minors' access to mobile apps. Neither has been in effect long enough to assess its effectiveness.

App Store Accountability Model

Utah, Louisiana, Texas, and Alabama require age verification, parent-linked accounts, and consent before every download.

This is the model SB237 follows.

Age-Signal Model

California and Colorado instead share an age signal with developers, who apply their own restrictions. No linked accounts, no per-download consent.

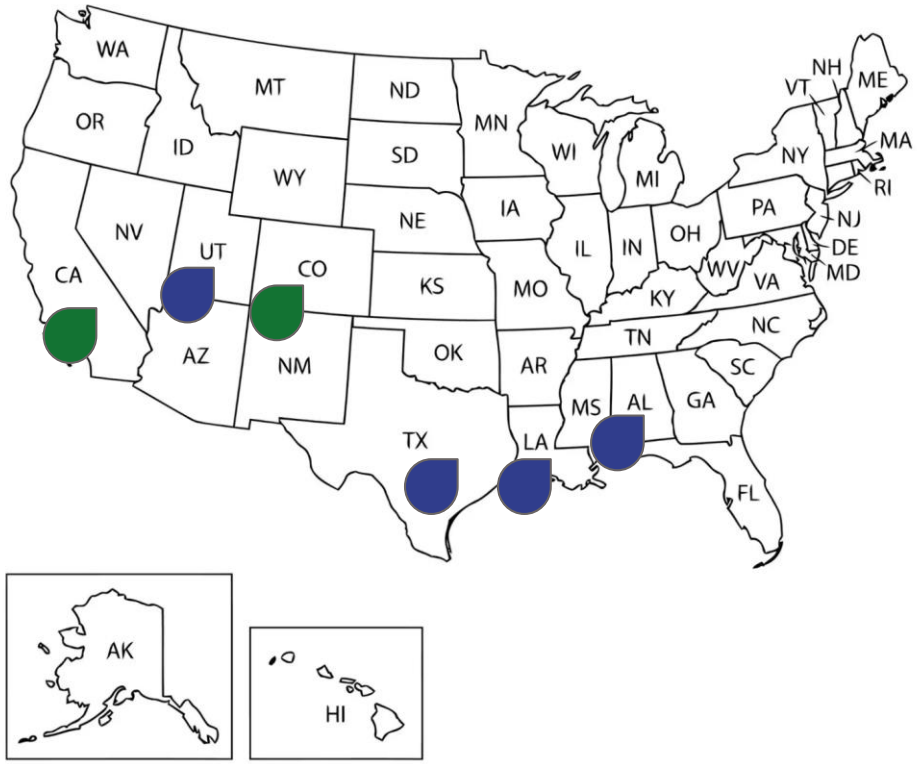
Current Policy Challenges

No evidence of success:

- Constant amendments
- Unsettled constitutionality

Key Takeaway

Every state is running the same "experiment" at once, and none of them can yet tell you if it's working.



JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Policy Recommendations

We recommend intermediary steps rather than a particular age assurance model: **strengthen existing tools**, then **build in evaluation.**"

1. MAKE EXISTING TOOLS TRUSTWORTHY

Require app store providers to **independently audit** developers' age ratings and content descriptions.

Require parental controls to be easy to **find, understand, and configure**.

Require app stores to provide a clear and conspicuous mechanism for users to **report inaccurate or misleading** app information.

2. TAKE INCREMENTAL STEPS & EVALUATE

Limit parental consent to apps **rated above the account holder's age category** and apps offering **in-app purchases**.

Require app store providers to give parents a **centralized dashboard** to set consent preferences by age rating or per download.

Require app stores providers to **collect and publicly report** data on verification, consent requests, blocked downloads, purchases, and complaints.

KEY TAKEAWAY

Strengthen existing protections, target requirements where risks are the greatest, and measure outcomes before adopting strict age assurance law.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Conclusions

01. THE EVIDENCE

The evidence supports the concerns that motivated the bill. Children face genuine online, privacy, and financial risks.

02. THE LIMITATION

Evidence does not yet show that app-store-level age verification and consent for every download will effectively prevent those harms.

03. DEPENDENCIES

The proposed framework depends on accurate age assessments, accurate app ratings, usable parental tools, meaningful consent, broad coverage, and strong privacy protections.

04. RECOMMENDATIONS

Focus on improving the reliability of existing tools, targeting the highest-risk situations, and requiring data that will allow policymakers to evaluate outcomes.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Thank You.

Questions & Discussion

SB237 - Bill Study Findings & Recommendations

Contact Information

Alejandro Velasco

William & Mary

Email: info@jcots.virginia.gov

Web: jcots.virginia.gov

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Overview

House Bill 707 - Dot-Gov Domains

Presented by Anna Dehmer, JCOTS William & Mary Intern

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Summary

House Bill 707 requires all government entities in the Commonwealth to utilize a dot-gov domain for their primary, public-facing, websites.

The goal of the bill is to reduce government website impersonation, increase public trust in government websites, and increase overall cybersecurity awareness throughout the commonwealth.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

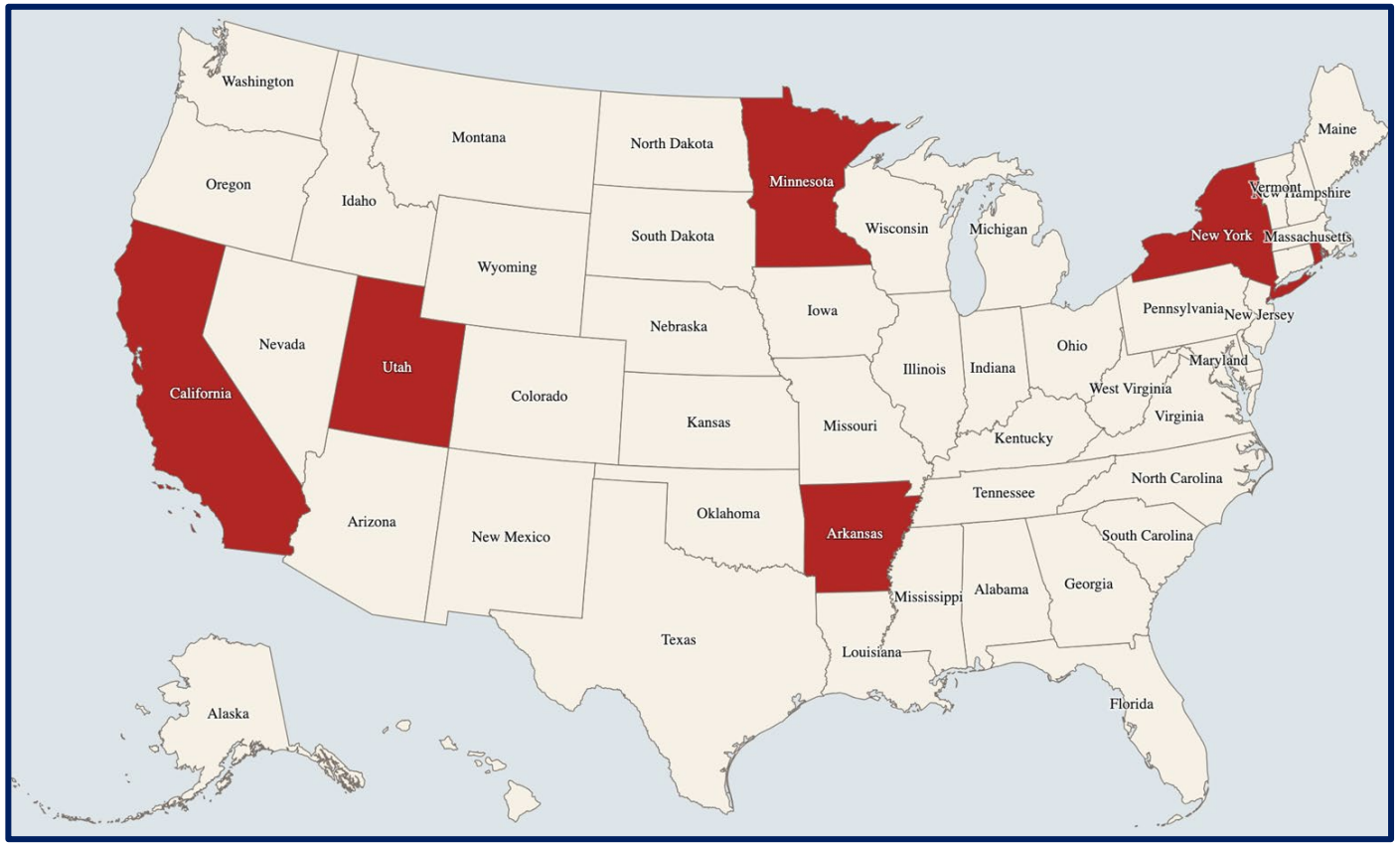
Dot-Gov Domains

- Only entities listed in the Census of Governments that go through a secure vetting process with CISA are eligible to receive a dot-gov domain
- Distributed by CISA (.gov) or VITA (.virginia.gov)
- Switching to the domain is free
- Agencies are recommended by CISA to keep their old domains active

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Other States' Legislation

- Arkansas
- Minnesota
- California
- New York
- Connecticut
- Utah



JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Federal Legislation

Requirement for all federal agencies to use a dot-gov domain

State and Local Cybersecurity Grant Program (SLCGP)

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Stakeholder Concerns

Dot-Edu and Dot-Museum Domains	Eligibility
Third Party Contractors	Implementation Time
Implementation Costs	Public Awareness

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Workgroup One Findings

- Many entities required to make the switch had no precedent for doing so elsewhere in the U.S.
- Requiring dot-gov may lessen the prestige carried by the domain name
- Many localities across the state have limited IT bandwidth, making transition difficult

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Workgroup Two Findings

- Libraries have the potential to distribute cybersecurity education to citizens across the Commonwealth
- VITA does not currently have the capabilities to enforce the bill
- Technology may change significantly in the upcoming years, rendering long timelines for implementation contradictory

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Primary Policy Questions

Cybersecurity

What is necessary to lessen website impersonation and improve cyber security education?

Eligibility

What classifies as a government agency? How can agencies know they are required to comply?

Funding and Enforcement

How can we support smaller entities making the transition? How can we make sure agencies are complying?

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Cybersecurity

Implement a cybersecurity education component to the bill, administered through the Library of Virginia

Require entities to list third party vendors that may be communicating directly with constituents on their website

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Eligibility

Only entities eligible for a dot-gov domain under the CISA definition are required to comply with the bill

If an entity is eligible for a dot-gov domain as well as a dot-edu domain or a dot-museum domain, they can choose to utilize a dot-edu or a dot-museum domain instead of a dot-gov domain

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Eligibility

VCU Health is exempt from the requirements of the bill

Add a provision stating if an agency became eligible following the bill's passage, they would have 3 years to transition

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Recommendations

Funding and Enforcement

- Introduce a budget amendment to assist small localities with the transition
- Promote the usage of the SLCGP funding
- Entities that fail to comply with the bill must submit a report to the GA

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

Thank you!

EVERRA HEALTH

Joint Commission on Technology and Science · Virginia General Assembly

Clinical AI Governance Starts Upstream.

Why traceable clinical interpretation infrastructure is the missing foundation of safe healthcare AI.

Noma Saeed · Founder & CEO, Everra Health · noma@everrahealth.com

August 2026

Richmond, Virginia

15-Minute Testimony

Views are my own.
Not legal or medical advice.

INTRODUCTION

Who I Am & Why I'm the Right Person to Speak to This

Software Engineer

15+ years building and operating within healthcare systems — I understand both the data architecture and the clinical workflow

Chief Operating Officer

COO of Dominion Heart and Vascular Center — ran alongside its President, interventional cardiologist Dr. Saquib Samee. Saw every angle: physician overwhelm, inconsistent patient reports, fragmented data.

Founder & CEO, Everra Health

Built the clinical interpretation infrastructure currently being evaluated at Dominion Heart and Vascular Center — 300 rules across 30 clinical domains

Patient

Diagnosed with a brain tumor three years ago — caused by chronic stress that physicians could not pinpoint as an underlying driver. The exact gap Everra was built to close.

WHY I'M HERE

Virginia is at a pivotal moment.

I spent 15 years inside Virginia healthcare — first as an engineer, then as an operator. I watched physicians get overwhelmed by data they couldn't synthesize in time. I watched patients leave without a clear picture of what was happening.

Then I became one of those patients.

The decisions this Commission makes will determine whether the next generation of clinical AI makes that better — or makes it invisible and unaccountable.

Views are my own. Not legal or medical advice.

WHY I BUILT THIS

I Saw the Gap From Every Angle. Then I Became a Patient.

As the COO

Running Dominion Heart and Vascular Center alongside its President — my husband, interventional cardiologist Dr. Saquib Samee — I watched physicians navigate 30 patients a day with fragmented data they couldn't synthesize in a 12-minute appointment. Patients left without a clear picture of what was happening. Physicians made decisions without the full story.

As the Operator

From the administrative side, I could see the structural problem clearly: the data existed — labs, wearables, EHR history, lifestyle signals. But no system was designed to structure it into a coherent clinical picture before the physician walked in. The interpretation bottleneck was invisible in the data. That made it ungovernable.

As the Patient

Three years ago, I was diagnosed with a brain tumor caused by chronic stress. Stress that my physicians could see signals of — but had no system to surface, structure, or flag as a clinical priority. By the time it became visible, the window had passed. I recovered. But I knew that gap was the problem I needed to solve.

Everra Health was built around a principle I believe should guide all clinical AI: that it be structured, traceable, and physician-reviewable. Whatever standards Virginia sets, I'd urge that those three principles be at the core.

THE PROBLEM

Healthcare AI Has a Governance Gap — and It Starts Before the Algorithm.

No standard

for clinical AI inputs

Unstructured Inputs

Most clinical AI ingests raw, fragmented data — unstructured lab values, free-text notes, disconnected wearable signals. There is no standard for what goes in.

No trace

from input to output

Black Box Outputs

When an AI tool makes a clinical recommendation, physicians often cannot see why. There is no audit trail, no reasoning chain, no accountability path when it is wrong.

No owner

of clinical AI error

No Liability Framework

When AI contributes to a clinical error, current law has no clear answer for who is responsible — the vendor, the health system, the physician, or the algorithm.

Virginia context:

Virginia health systems are deploying clinical AI now. Without a traceability standard, the Commonwealth has no mechanism to audit a clinical AI decision, protect a patient harmed by one, or hold any party accountable.

THE CORE ISSUE

The Real Risk Is What Happens Before the AI Makes a Decision.

Consider a bridge.

We require engineers to document every structural assumption before a bridge is built. If it fails, we can trace exactly what went wrong and who decided it. Clinical AI has no equivalent requirement. When it fails, there is nothing to audit.

Current State — No Traceability	Required State — Structured & Traceable
Raw data enters AI with no standardized structure	Data structured against clinical guidelines before AI
Clinical reasoning is opaque — no audit trail	Every rule, signal, and output is documented
Physician cannot explain why AI flagged a risk	Physician sees full reasoning chain, can override
Error occurs → no chain of accountability	Error occurs → exact point of failure is auditable
Regulator has nothing to review	Regulator can review deterministic logic trail

THE ENGINE

How Structured Clinical Reasoning Works: The Engine That Makes Governance Possible

1

Patient Signals

Labs, wearables, EHR inputs — cardiometabolic, inflammatory, metabolic, sleep

2

Clinical Rules Applied

300 rules / 30 domains mapped against ACC/AHA, ESC, ADA guidelines

3

L1/L2/L3 Output

Each signal classified — every classification has a documented reasoning chain

4

Priority Picture Built

Signals synthesised — root cause, trend, confidence score

5

Physician Reviews

Structured output delivered before the appointment — physician decides, can override

Why this matters for governance:

Every step is documented. If the output is wrong, a regulator can trace exactly which rule fired, which signal triggered it, which guideline it was anchored to, and what the physician was shown. That is what auditable clinical AI requires — and what most systems today cannot provide.

EVIDENCE & SOLUTION

Four Properties That Make Clinical AI Governable

Everra Health has built each of these into its architecture. Together, they form a model Virginia can reference when setting standards.

01

300 rules · 30 domains

Deterministic Rules, Not Black Box

300 clinical rules across 30 domains. Every rule is explicit, documented, and anchored to peer-reviewed clinical guidelines (ACC/AHA, ESC, ADA), with the source guideline version recorded for each rule. A regulator can read exactly what the system did and why.

02

L1 · L2 · L3 documented

3-Tier Severity Output with Full Audit Trail

Every output is classified L1, L2, or L3 severity. Each classification has a documented reasoning chain — which signals triggered it, what guidelines were applied, what the physician was shown.

03

Physician override always available

Physician Stays in Control — Architecturally

The system structures the clinical picture and delivers it to the physician before a decision is made. It does not decide. The physician reviews, can override, and remains the accountable party.

04

Peer-reviewed · Dr. Babar Ali, MD

Independent Clinical Validation at Rule Level

Rule sets were reviewed by an independent practicing cardiologist with no financial relationship to Everra. Validation is documented at the individual rule level — the standard Virginia should require.

EVIDENCE

What the Engine Surfaces. And Why That Matters for Governance.

3

reviewing cardiologists

Independent clinical peer review panel evaluating rule logic, citations, and thresholds

Dr. Babar Ali · Dr. Atique Mirza · Dr. Mohsin Ijaz

150+

patients processed

Patient datasets structured and interpreted through the engine during evaluation

300

rules · 30 domains

Clinical rules validated against peer-reviewed guidelines, reviewed by independent cardiologist

The engine is being evaluated within Dominion Heart and Vascular Center's cardiology workflow. In internal physician review, the structured output has surfaced cardiometabolic patterns that a physician would not typically synthesize in a 12-minute appointment — patterns the reviewing cardiologist judged clinically meaningful.

— Dr. Saquib Samee, Interventional Cardiologist & Clinical Co-Founder · Dominion Heart and Vascular Center

The governance principle: if the structured picture had been wrong, every step of the reasoning would be auditable. That auditability — not the engine itself — is what I'd urge Virginia to require of all clinical AI.

VIRGINIA CONTEXT

Why This Matters for Virginia — and Why Now.

Virginia Health Systems Are Deploying Clinical AI Now

Active deployment

Major Virginia health systems — including Inova, VCU Health, Sentara, and UVA Health — are actively integrating AI tools into clinical workflows without any state-level traceability or accountability standard.

Virginia Patients Have No Recourse When Clinical AI Fails

Patient protection

Under current Virginia law, when an AI-assisted clinical decision contributes to patient harm, there is no clear liability framework, no mandatory audit trail, and no disclosure requirement.

Virginia Can Set the National Standard

Policy leadership

Virginia has led on technology policy before. The EU AI Act has established traceability requirements for high-risk AI in Europe. Virginia has an opportunity to establish the domestic equivalent for clinical settings.

The EU AI Act — Coming to U.S. Clinical Practice

Regulatory alignment

The EU AI Act classifies clinical decision support as high-risk AI; under its 2026 amendments, traceability and human-oversight duties phase in through 2028. Vendors already build to them — Virginia can align proactively.

POLICY RECOMMENDATIONS

What Virginia Should Require of Clinical AI Systems

1

Require Traceability Standards for Clinical AI

Priority ask

Any AI system used in clinical decision-making in Virginia should maintain a complete, auditable reasoning trail — from input data to output recommendation. Physicians and regulators must be able to review exactly what the system did and why.

2

Mandate Clinician-in-the-Loop Architecture

Patient safety

Virginia should require that clinical AI systems are architecturally designed so that physician judgment is central and cannot be bypassed. The AI structures — the physician decides. This must be law, not a vendor promise.

3

Establish Independent Clinical Validation Requirements

Accountability

AI clinical rule sets should undergo peer review by independent licensed clinicians before deployment. Validation must be documented at the rule level, not just system-wide, and available to regulators on request.

4

Define Liability Framework for Clinical AI Errors

Legal framework

Virginia should initiate a study to define who bears liability when clinical AI contributes to patient harm. This framework should rest on the principle: traceable AI can be audited; untraceable AI cannot be defended.

EVERRA HEALTH

Virginia Can Lead on Clinical AI Governance.

Every patient in Virginia deserves to know why an AI flagged their health — and every regulator deserves the trail to audit it.

What success looks like:

- ✓ A Virginia patient can ask: why did the AI flag this? — and get a documented answer
- ✓ A Virginia regulator can audit any clinical AI decision in a licensed health system
- ✓ A Virginia physician is never bypassed by an algorithm — their judgment is always central
- ✓ A Virginia court has a legal framework for assigning liability when clinical AI causes harm

Thank you to Chair and members of the Commission for your time and this opportunity.

Noma Saeed · noma@everrahealth.com · www.everrahealth.com