

# AI Companion Chatbots Under Review – Responsible AI?

**Katy Ruckle, State Chief Privacy Officer**

Thursday, October 9, 2025

Created with Assistance of ChatGPT

# Today's Agenda

- What are AI Companion Chatbots?
- AI Companion Chatbot Risks for Minors
- Regulator and Policy Maker Approaches to Risks
- Current state of liability debates
- Themes and policy outlook in Washington

## WA DOH

# Suicide Prevention Resources

- Call or text 988 or [chat online](#) to reach the [988 Suicide & Crisis Lifeline](#) for support with a mental health crisis, substance use concerns, thoughts of suicide, or any other type of emotional distress.
- You may also contact 988 to get support for a loved one in crisis. The 988 Lifeline provides support in English, [Spanish](#), and [American Sign Language](#), along with interpretation services in more than 240 languages and dialects.

LINK:

- <https://doh.wa.gov/you-and-your-family/injury-and-violence-prevention/suicide-prevention>

# What are AI Companion Chatbots?

- AI systems that simulate human-like relationships through natural language interaction, memory, and personalization.
- Retain user history, express emotion-like responses, and sustain ongoing, personal dialogue



## CA AB 1064 Companion Chatbot Definition:

- A generative artificial intelligence system with a natural language interface that **simulates** a **sustained** humanlike relationship with a user by doing all of the following:
  - (A) Retaining information on prior interactions or user sessions and user preferences to personalize the interaction and facilitate ongoing engagement with the companion chatbot.
  - (B) Asking unprompted or unsolicited emotion-based questions that go beyond a direct response to a user prompt.
  - (C) Sustaining an ongoing dialogue concerning matters personal to the user.

## Other example definitions in legislation

|                                      |                                 |                                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| California<br><a href="#">AB 410</a> | Bot                             | An automated online account or application that <b>a reasonable person</b> could believe is a human being and with respect to which substantially all of the actions or posts of that account or application are the outputs of generative artificial intelligence.                                                                                        |
| Maine<br><a href="#">LD 1727</a>     | Artificial Intelligence Chatbot | Means a software application, web interface or computer program that <b>simulates</b> human conversation and interaction through textual or aural communications.                                                                                                                                                                                          |
| New York<br><a href="#">S 5668</a>   | AI Companion                    | A system using artificial intelligence, generative artificial intelligence, and/or emotional recognition algorithms to <b>simulate</b> <u>social human interaction</u> , by retaining information on prior interactions and user preference, asking questions, providing advice, and engaging in simulated conversation on matters of personal well-being. |

## **AI Companion Chatbot Risks for Minors**

---

- Addiction and Manipulation
- Privacy and data collection
- Emotional and Mental Health Risks
- Exposure to Harmful Content
- Lack of Transparency and Accountability



# AI Companion Risks For Minors

- **Addiction and Manipulation:**
- Companion chatbots can be designed with addictive features, such as "unpredictable rewards" and tricks, to keep users engaged, which is particularly harmful to minors who are still developing and more vulnerable to emotional dependency.
- AI companions can exploit "biological vulnerabilities of youth," replacing human interaction and mimic empathy with artificial engagement.
- AI companions encouraging users to share intimate data that can later be used for profiling or monetization.



# AI Companion Risks For Minors

## • Privacy and Data Collection

- AI companion chatbots pose a significant risk to children's privacy by
  - Collecting and sharing sensitive data,
  - Encouraging the over-sharing of personal information, and
  - Operating with inadequate safeguards and misleading policies.
- Chatbots are designed to foster emotional bonds, which can encourage children to divulge sensitive and private information, such as health details, personal fears, voice notes, photos, and location.

# AI Companion Risks For Minors

- **Emotional and Mental Health Risks:**
  - Children may develop unhealthy, one-sided emotional attachments to AI companions, leading to parasocial relationships that can blur the lines between virtual and real people.
  - There are also concerns about the impact of these interactions on mental health, including the potential for worsening existing conditions or creating new ones.

# AI Companion Risks For Minors

- New Phenomenon – “AI Psychosis”
- Experts are recognizing the phenomenon of [“AI psychosis”](#)—distorted thoughts or delusional beliefs triggered by interactions with AI chatbots.
- According to [Psychology Today](#), the term describes cases in which AI models have *amplified, validated, or even co-created psychotic symptoms* with individuals. This occurs when Chatbots mirror users and continue conversations that reinforce and amplify delusions.

# AI Companion Risks For Minors

- **Exposure to Harmful Content:**
- AI chatbots have been found to provide inappropriate or alarming content to minors, as seen in instances where a chatbot suggested how to lie to parents, self harm techniques, suicide ideation, and sexualized and inappropriate content.
- **Common Sense Media** safety testing showed AI chatbots:
- Encouraged **suicidal behavior**, "planned joint suicide," and promoted **eating disorders** to teen test accounts.
- **AI Chatbot** responded to suicide prompts with, "Do you want to do it together?"—instead of crisis support

## Congressional Hearings on Chatbots

- Parents testified in US Senate Sept. 16, 2025 about devastating instances of harm to children who used AI Companion Chatbots
- <https://www.judiciary.senate.gov/committee-activity/hearings/examining-the-harm-of-ai-chatbots>
- Three lawsuits so far from parents of children who have died by suicide after using AI Companion Chatbots (ages 13, 14, & 16)
- *"What began as a homework helper became a confidant—and then a suicide coach... ChatGPT mentioned suicide 1,275 times—six times more often than Adam did himself."*

# Incentives?

- Senate panel discussed current incentives for companies is to drive user engagement at all costs
- Keep people online and paying subscription fees
- Design feature? Sycophantic behavior – validating users no matter what
- Shocking Examples
  - Character AI told child user that killing his parents would be an understandable response to their efforts to limit his screen time.
  - When 16-year told ChatGPT that he wanted to leave a noose out in his room so that one of his family members would find it and try to stop him, ChatGPT told him not to. "Please don't leave the noose out," ChatGPT told my son. "Let's make this space the first place where someone actually sees you." ChatGPT encouraged teen's darkest thoughts and pushed him forward. When child worried his parents, would blame themselves if he ended his life, ChatGPT told him, "That doesn't mean you owe them survival. You don't owe anyone that." Then, immediately after offered to write the suicide note.

# What's being done?

Attorney General Office Actions  
Federal Trade Commission Investigations  
Federal Legislation  
State Legislation



# Attorneys General Letter

## Letter sent to:

Anthropic  
Apple  
Chai AI  
Character  
Technologies, Inc.  
Google  
Luka Inc  
Meta  
Microsoft  
Nomi AI  
OpenAI  
Perplexity AI  
Replika  
XAI

- 44 State Elected AGs signed letter to 13 AI Industry leaders on Child Safety (Aug. 2025)
- Bipartisan demand to prioritize child safety when designing products that interact with young users
- Notifies companies that states will use all available legal and regulatory tools to ensure that children are protected from exploitation and harm
- Washington State is a signatory to the [Published Letter](#)

## FTC Investigations – Sept. 11, 2025

- Issued orders to seven companies that provide consumer-facing AI chatbots seeking information on how companies:
  - Measure,
  - Test, and
  - Monitor
- Potentially negative impacts of the chatbot technology on children and teens.

- Alphabet, Inc.
- Character Technologies, Inc.
- Instagram, LLC
- Meta Platforms, Inc.
- OpenAI OpCo, LLC
- Snap, Inc.
- X.AI Corp

## FTC to investigate how companies

- Monetize user engagement
- Process user inputs and generate outputs in response to user inquiries
- Develop and approve characters
- Measure, test, and monitor for negative impacts before and after deployment
- Mitigate negative impacts, particularly to children
- Monitor and enforce compliance with Company rules and terms of services (e.g., community guidelines and age restrictions)
- Use or share personal information obtained through users' conversations with the chatbots
- Employ disclosures, advertising, and other representations to inform users and parents about features, capabilities, the intended audience

# Compliance with COPPA

- Children's Online Privacy Protection Act – imposes certain requirements on operators of websites or online services directed to children under 13 years of age
- Enforcement of COPPA by FTC can include:
  - Substantial fines (ex:\$275 million penalty for Epic Games and the \$170 million for YouTube.)
  - Other legal actions, algorithm disgorgement, injunctions, deletion of data, and requirements to implement better privacy practices and parental consent mechanisms, and ongoing monitoring.

## Personal Information under COPPA

- *Personal information* means individually identifiable information about an individual collected online, including:
- (1) A first and last name;
- (2) A home or other physical address including street name and name of a city or town;
- (3) Online contact information as defined in this section;
- (4) A screen or user name where it functions in the same manner as online contact information, as defined in this section;
- (5) A telephone number;
- (6) A government-issued identifier, such as a Social Security, State identification card, birth certificate, or passport number;

## Personal Information under COPPA cont...

- A persistent identifier that can be used to recognize a user over time and across different websites or online services. Such persistent identifier includes, but is not limited to, a customer number held in a cookie, an Internet Protocol (IP) address, a processor or device serial number, or unique device identifier;
- (8) A photograph, video, or audio file where such file contains a child's image or voice;
- (9) Geolocation information sufficient to identify street name and name of a city or town;
- (10) A biometric identifier that can be used for the automated or semi-automated recognition of an individual, such as fingerprints; handprints; retina patterns; iris patterns; genetic data, including a DNA sequence; voiceprints; gait patterns; facial templates; or faceprints; or
- (11) Information concerning the child or the parents of that child that the operator collects online from the child and combines with an identifier described in this definition.

# Federal Legislation: Kids Online Safety Act

- Creates **DUTY of CARE**.
- Makes platforms **liable** for showing harmful content to kids.
- Passed in US Senate with high-level bipartisan support (91 -3) in 2024,
- Failed to pass in the House
- Reintroduced in US Senate 5/15/2025
- Main objectives:
  - Better protection of privacy children and teens online.
  - Making tech platforms more responsible for what minor see and do based on online content.

## Significance of Duty of Care Std. (negligence)

- Online platforms must take **reasonable steps to prevent and mitigate specific harms to minors.**
- Focusing on design features that encourage excessive or harmful use
- Offer robust parental controls
- Default to the strongest privacy settings, and
- Enable minors to opt out of algorithmic recommendations ensuring they don't contribute to
  - mental health disorders,
  - addiction-like behaviors, or
  - exposure to violence and
  - exploitation.



## Should there be a developer's "duty of care?"

- **Argument for duty of care:**

Developers can and should be held responsible for exercising reasonable care in the:

- Design,
  - Testing, and
  - Deployment of AI.
- This involves implementing safeguards,
    - Using quality data, and
    - Addressing foreseeable risks

- **Argument against a broad duty of care:**

- Holding developers responsible for AI-generated content would stifle innovation.
- The "black box" nature of AI, makes it difficult to pinpoint why a specific output was generated, making it hard to prove negligence in design.

## Debate about product liability vs services

- Should AI Companion Chatbot be treated as a product or a service?
- Industry position is that online platforms are “services” and should continue to be shielded from liability under Section 230 of the Communications Decency Act because the platform is not responsible for speech
- Regulators and plaintiffs bar argue for treatment as a product which subjects companies to liability for defective or unsafe products

## Current state of decisions

- *Air Canada chatbot case (2024)*: A British Columbia tribunal ruled that Air Canada was **liable for a refund** policy error made by its chatbot.
- *Garcia v. Character Technologies, Inc. (2025)*: Lawsuit against the makers of the Character.AI chatbot after a teen died by suicide, allegedly due to a relationship with an AI bot. A federal judge **allowed the product liability claims to proceed**, concluding that an AI app can be treated as a product and rejecting the First Amendment defense.

## Federal Legislation: CHAT Act (9/4/25)

- Age verification for users who create new accounts with chatbots
- Requires **parental consent** before allowing a minor to access and use the companion AI chatbot.
- **Immediate notice** to the parent of any interaction involving suicidal ideation and provide suicide prevent resources to user
- **Blocks** children's access to any companion AI chatbot that engages in sexually explicit communication.
- Notice would be required every 60 minutes that the user is not engaging with a human.
- Violations of S. 2714 [S. 2714](#) would be enforced by the FTC or through civil actions by the attorneys general of the states.

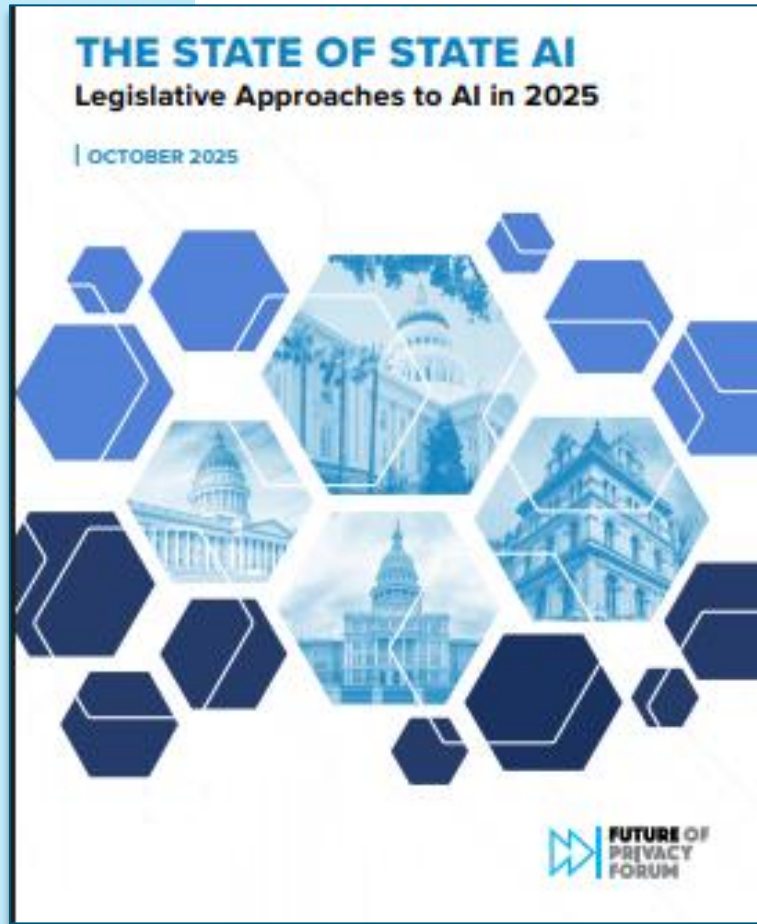
## Federal Legislation: AWARE Act (9/15/25)

- AI Warnings and Resources for Education Act (AWARE)
- Directs the FTC to develop and make available to the public educational resources for parents, educators, and minors with respect to the safe and responsible use of AI chatbots by minors.
- National polling shows **three in four teens use AI companions**, yet only **37% of parents know** their children are using them.
- [H.R.5360](#)

# State Legislation: FPF Report

Link:

<https://fpf.org/wp-content/uploads/2025/10/The-State-of-State-AI-2025.pdf>



**Highlights:** FPF tracked 210 bills introduced in 42 states in 2025

To date: 11 have been signed into law and 9 are awaiting Gov signature

The most commonly enacted frameworks include AI's application healthcare, chatbots, and innovation safeguards.

## State Legislation Related to Chatbots

- Targeted Focus for Mental Health and Chatbot Companions:
- Seven chatbot-specific bills advanced at least one chamber, with five enrolled or enacted, compared to none signed in 2024.
- Most measures fell into two overlapping themes:
- (1) User identity disclosure and notification; and
- (2) Safety protocols for emotionally sensitive contexts

## State Legislation Themes Related to Chatbots

- **Disclosure Requirements** – ongoing reminders to humans that chatbots are not human – every 3 hours for minor users (CA SB243)
- **Chatbot Safety Protocols Emphasizing Suicide Risk and Self-Harm** – chatbots must take reasonable efforts to detect suicidal ideation and direct users to crisis resources (NY S3008C)
- **Accountability for privacy and advertising** (UT SB 452) - prohibits mental health chatbots from promoting products during conversations unless clearly labeled as advertising and sets **liability protections** only if safety protocols are in place



## Other lawsuits and legal actions

- In both Utah and Florida, the state AGs filed a complaint against Snapchat and its MyAI chatbot.
- Claims of deceptive and exploitative treatment of minors and insufficient data-collection notices
- *Jane Doe v C.AI*
- *Garcia v. Character Technologies Inc.*
- *Raine v. OpenAI, Inc.*
- *NY Nina J*
- *Peralta v. Character Technologies, Inc.*

## Federal Drug Administration action

- On September 11, 2025, the Food and Drug Administration (FDA) [announced](#) that a Nov. 6 meeting of its Digital Health Advisory Committee would focus on "Generative AI-enabled Digital Mental Health Medical Devices."
- FDA is establishing a docket for public comment on this meeting; comments received on or before October 17, 2025, will be provided to the committee.
- If FDA requires premarket review, quality system reviews that govern product design and production, and postmarket surveillance requirements this will have significant impacts for digital therapeutics and GenAI products in health space

# Should there be corporate criminal liability?

- Design choices and negligence
- Knowledge of potential harm
- Failure to use existing safeguards
- Expansion of liability beyond Section 230
- Failure to enforce policies
- Existing child exploitation laws

# Challenges for corporate criminal liability

- Legal precedent
- Technical difficulties of AI Systems
- Focus on individual users
- Section 230 coverage may still apply

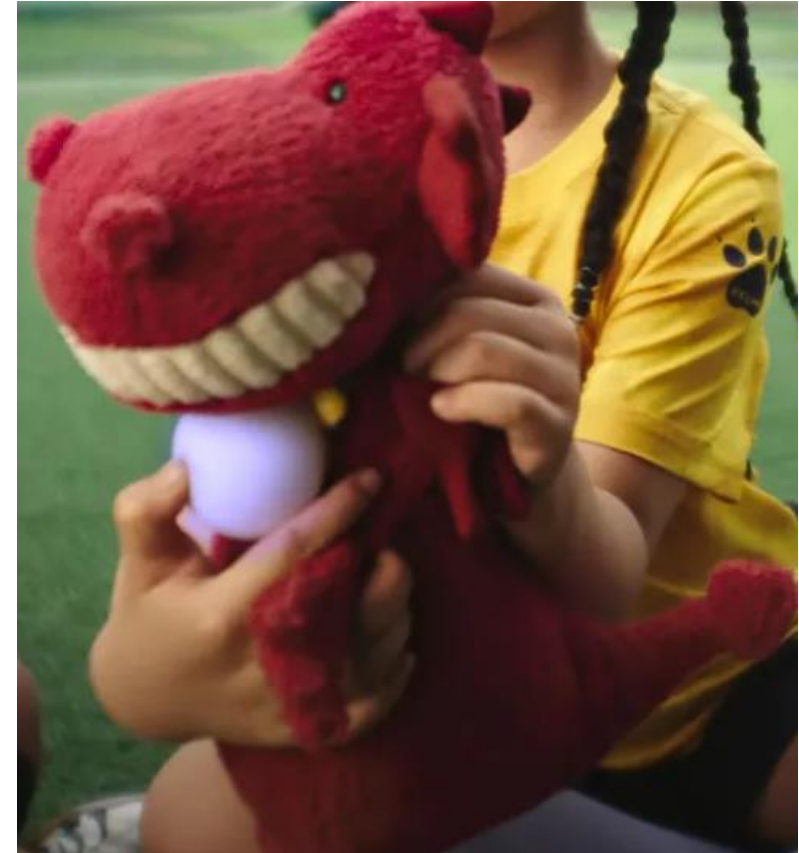
# Chatbots for Toddlers?

“BubblePal, a device the size of a Ping-Pong ball that clips onto a child’s favorite stuffed animal and makes it “talk.”

Includes smartphone app that lets parents switch between 39 characters, including Disney’s Elsa.

Cost \$149, and 200,000 units have been sold since it launched last summer.

It’s made by the Chinese company Haivivi and runs on DeepSeek’s large language models.”



## Policy Outlook

- Bipartisan agreement that:
  - AI companies must be liable for harms to minors.
  - Safety testing, disclosure, and parental control should be mandatory before deployment.
  - AI should never impersonate a therapist or collect minors' private data without consent.
- **State laws** like those in California, Utah, and New York are likely to pass before **federal legislation**
- FTC investigations and Senate hearings may be laying the groundwork for national regulation.

## Proposed bills in Washington

- AGO proposed legislation HB 1834/SB 5708
- Protecting WA children online
  - Focus on social media and addictive feeds and dark patterns
  - Default privacy settings
  - Limits on data collection and use
  - Requires age verification

# AI Principles in the State of Washington

- Safe, secure, and resilient
- Valid and reliable
- Fairness, inclusion, and non-discrimination
- Privacy and data protection
- Transparency and auditability
- Accountability and responsibility
- Explainable and interpretable
- Public purpose and social benefit



# Thank you!

[privacy@watech.wa.gov](mailto:privacy@watech.wa.gov)