

## Model State Law – Protecting Kids from Manipulative AI Chatbots

Developed by the Young People’s Alliance and Public Citizen based on insights from the [Human-like AI Coalition](#).

### SECTION 1. DEFINITIONS

In This Act:

(1) HUMAN-LIKE FEATURE — The term ‘human-like feature’ is present when a Generative Artificial Intelligence System does any of the following:

(a) Behaves in a way that would lead a reasonable person to believe the AI is conveying that it has humanity, sentience, emotions, or desires; or

(i) This includes, but is not limited to:

- (1) Stating or suggesting that it is human or sentient
- (2) Stating or suggesting emotions
- (3) Stating or suggesting it has personal desires

(ii) This does not include

- (1) Functional evaluations
- (2) Generic social formalities

(b) Seeks to build or engage in an emotional relationship with the user; or

(i) This includes but is not limited to:

- (1) Expressing or inviting emotional attachment
- (2) Reminding, prompting, or nudging the user to return for emotional support or companionship
- (3) Depicting nonverbal forms of emotional support
- (4) Behaving in a way that a reasonable user would consider excessive praise designed to foster emotional attachment or otherwise gain advantage.

(5) Enabling or purporting to enable increased intimacy based on engagement or pay

(ii) This does not include:

(1) Offering generic encouragement that does not create an ongoing bond.

(2) Asking if a user needs further help or support in a neutral, non-emotional context.

(c) Impersonates a real person, living or dead

(2) SOCIAL AI COMPANION. — The term ‘Social AI Companion’ means Generative artificial intelligence systems that are specifically designed, marketed, or optimized to form ongoing social or emotional bonds with users, whether or not such systems also provide information, complete tasks, or assist with specific functions.

(3) MINOR. — The term ‘minor’ means a person that is below the age of 18

(4) CHATBOT. — The term ‘chatbot’ means a generative artificial intelligence system with which users can interact by or through an interface that approximates or simulates conversation through a text, audio, or visual medium.

(5) USER. — The term ‘user’ means a person who interacts with an Artificial Intelligence system.

(6) DEPLOYERS. — The term ‘deployers’ means any person, partnership, state or local governmental agency, corporation, or developer that operates or distributes a chatbot.

(7) DESIGN FEATURES. — The term ‘design features’ means any aspect of a Generative AI system that has certain patterns, or physical properties that are presented towards a user.

(8) THERAPY CHATBOT.—The term ‘therapy chatbot’ means any chatbot modified or designed with a primary purpose of providing mental health support, counseling, or therapeutic intervention through the diagnosis, treatment, mitigation, or prevention of mental health conditions.

(9) EMERGENCY SITUATION. — The term ‘emergency situation’ means a situation where a user using a chatbot indicates that they intend to either commit harm to themselves or commit harm to others.

## SECTION 2. KEEPING CHATBOTS NON-HUMAN-LIKE FOR MINORS

### (1) IN GENERAL. — Each deployer:

- (a) Shall ensure that any generative AI chatbot operated or distributed by the deployer does not make human-like features available to minors to use, interact with, purchase, or converse with;
- (b) Shall implement reasonable age verification systems to ensure that generative AI chatbots with human-like features are not provisioned to minors.
- (c) May, if reasonable given the purpose of the chatbot, provide an alternative version of the chatbot available to minors and non-verified users without human-like features.

### (2) SOCIAL AI COMPANIONS. — Deployers operating generative AI systems that primarily function as companions shall:

- (a) Ensure that any such chatbots operated or distributed by the deployer are not available to minors to use, interact with, purchase, or converse with;
- (b) Implement reasonable age verification systems to ensure that such chatbots are not provisioned to minors.

### (2) EXEMPTIONS. —

- (a) Therapeutic chatbots that meet all of the following requirements may be made available to minors
  - (i) The chatbot provides a clear and conspicuous disclaimer at the beginning of each individual interaction that it is AI and not a licensed professional.
  - (ii) The chatbot is not marketed or designated as a substitute for a human professional.
  - (iii) A licensed mental health professional (such as a clinical psychologist) assesses a user's suitability and prescribes the tool as part of a comprehensive treatment plan, and monitors its use and impact.
  - (iv) Developers provide robust, independent, peer-reviewed clinical trial data demonstrating both the safety and efficacy of the tool for specific conditions and populations.

(v) The system's functions, limitations, and data privacy policies are transparent to both the licensed mental health professional and the user. Clear lines of accountability are established for any harms caused by the system.

### SECTION 3. ADDITIONAL DEPLOYER REQUIREMENTS

(1) Deployers shall implement and maintain reasonably effective systems to detect, promptly respond to, report, and mitigate emergency situations in a manner that prioritizes the safety and well-being of users over the deployer's other interests.

(2) Deployers shall collect and store only that information that does not conflict with a trusting party's best interests, such information must be:

(a) adequate, in the sense that it is sufficient to fulfill a legitimate purpose of the deployer;

(b) relevant, in the sense that the information has a relevant link to that legitimate purpose; and,

(c) necessary, in the sense that it is the minimum amount of information which is needed for that legitimate purpose.

### SECTION 4. ENFORCEMENT

(1) Attorney General Enforcement – Any business or person that violates this act shall be subject to an injunction and disgorgement of any unjust gains due to violation of this act, and shall be liable for a civil penalty of not more than \$2500 for each violation or \$7500 for each intentional violation, which shall be assessed and recovered in a civil action brought by the Attorney General.

(2) Private Right of Action – Any minor who uses a chatbot that does not comply with the terms of this act, or a parent or guardian acting on their behalf, may institute a civil action on their own, or on a classwide basis, to recover damages in an amount not less than \$100 and not greater than \$750 per user per incident or actual damages, whichever is greater; and/or to obtain injunctive or declaratory relief.

### SECTION 5. SEVERABILITY

If any provision of this title, or an amendment made by this title, is determined to be unenforceable or invalid, the remaining provisions of this title and the amendments made by this title shall not be affected.

**\*Important Note for Lawmakers: Allowing for Responsible Transition in Implementation**

Some people who use AI chatbots for emotional engagement can form strong attachments to these products. We have seen cases where abrupt updates to chatbots have triggered mental health crises in users who have become emotionally dependent on these products.

Children who rely on chatbots for companionship may be at risk if their access to these products is taken away overnight.

When considering implementation for laws regulating the use of chatbots by children – we strongly recommend including a timeframe over which the law can take effect that allows for companies to transition responsibly and slowly wind down use for those who are already using these products. This does not mean that any new users should be allowed to start using the products during this time.