AMENDMENT # 1

PARENTS & KIDS SAFE AI ACT

SEC. 1. Title.

This measure shall be known, and may be cited, as the Parents & Kids Safe AI Act.

SEC. 2. Findings & Declarations and Statement of Purpose.

The People of the State of California find and declare the following:

A. Any use of artificial intelligence (AI) by children under 18 must be safe and allow parents to monitor and limit their child's use of AI.

B. It is essential that California sets clear, strong, and enforceable standards that protect kids and give parents tools to protect their children, while encouraging new innovations that benefit schools and families and strengthen California's economy. This strong, responsible, and balanced approach protects children and families and supports economic progress that helps our state thrive.

C. To regulate the fast-growing AI industry, California needs standards that can keep pace with today's technology. Updating state law by requiring AI companies to use readily available tools — such as age estimation and age-appropriate content filters — to better protect kids will bring the state's rules in line with where technology is now, keeping children safe and parents involved and aware.

D. Therefore, the People of the State of California enact this measure in furtherance of the findings and declarations stated herein to:

(1) Require AI companies to use age assurance technology to distinguish children from adults, and when age cannot be determined, default to protective safeguards;

(2) Prohibit AI companies from targeting advertising to children or selling children's personal data without parental consent;

(3) Require AI companies to identify and develop safeguards to prevent generating harmful content for children, such as sexually explicit content and content promoting suicide, self-harm, or violence;

(4) Prohibit deceptive or manipulative design of AI systems used by minors that would create emotional dependence, simulate romantic relationships, or make child users think they are talking to a human;

(5) Require AI companies to create easy-to-use tools that give parents the choice and ability to monitor and limit their child's use of AI technology;

(6) Require AI companies to provide the ability for parents to be notified if a child expresses an intent to harm themselves;

(7) Require AI companies to publish and make public, for the benefit of parents and the public, their child safety policies and information about their parental controls;

(8) Require the Attorney General to enforce the child safety requirements and impose penalties on AI companies; and

(9) Require independent audits reviewing AI companies' compliance with the child safety requirements and the submission of such reports to the state.

SEC. 3. Division 8, Chapter 22.6 (commencing with Section 22601) of the Business and Professions Code, is amended (text deleted is denoted in ~~strikeout type~~ and text added is denoted in *underlined and italicized type*) to read:

Chapter 22.6. ~~Companion Chatbots~~ *Parents & Kids Safe AI Act*

22601. As used in this chapter:

(a) "*Artificial Intelligence System* ~~Artificial intelligence~~" means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

(b) (1) "Companion chatbot" means an artificial intelligence system with a natural language interface that provides adaptive, human-like responses to user inputs and is capable of meeting a user's social needs, including by exhibiting anthropomorphic features and being able to sustain a relationship across multiple interactions.

(2) "Companion chatbot" does not include any of the following:

(A) A bot that is used only for customer service, a business' operational purposes, productivity and analysis related to source information, internal research, or technical assistance.

(B) A bot that is a feature of a video game and is limited to replies related to the video game that cannot discuss topics related to mental health, self-harm, sexually explicit conduct, or maintain a dialogue on other topics unrelated to the video game.

(C) A stand-alone consumer electronic device that functions as a speaker and voice command interface, acts as a voice-activated virtual assistant, and does not sustain a relationship across multiple interactions or generate outputs that are likely to elicit emotional responses in the user.

~~(c) "Companion chatbot platform" means a platform that allows a user to engage with companion chatbots.~~

(c) (1) "Covered Artificial Intelligence (AI) System" means an artificial intelligence software application, web interface, companion chatbot, or computer program that is accessible to the general public in the state and that primarily simulates human conversation and interaction through textual, visual, or aural communications and is capable of generating contextually adaptive responses across multiple interactions, to non-business consumers.

(2) "Covered Artificial Intelligence (AI) System" does not include an artificial intelligence software application, web interface, companion chatbot, or computer program that is any of the following:

(A) Designed to provide outputs relating to a narrow and discrete topic;

(B) Used solely for commercial use by business entities;

(C) A speaker and voice command interface or voice-activated virtual assistant for a consumer electronic device; or

(D) Used by a business solely for internal purposes.

(d) "Child" means an individual under the age of 18.

(e) "Child Targeted Advertising" means "Cross-context behavioral advertising" directed at children as that term is defined in Section 1798.140(k) of the California Civil Code.

(f) "Child Safety Policy" means a public-facing document describing protective measures taken by a Covered AI System to mitigate identified Child Safety Risks.

(g) "Child Safety Risks" means reasonably foreseeable risks of severe harm to a child.

(h) "Encrypted User Content" means content (including audio, visual, or textual content) that is stored, transmitted, or held in a manner that is end-to-end encrypted or otherwise cryptographically protected such that the Provider of a Covered AI System cannot access the cleartext content without notice to the user or customer. For the avoidance of doubt, nothing in this chapter shall be construed to require a Provider of Covered AI System to alter, weaken, bypass, or otherwise modify its cryptographic or security protections in order to access or disclose the cleartext of Encrypted User Content.

(*i*~~d~~) "Office" means the Office of Suicide Prevention established pursuant to Section 131300 of the Health and Safety Code.

(*j*~~e~~) "Operator" means a person who makes a companion chatbot ~~platform~~ available to a user in

the state.

(k) "Parent" means a parent or legal guardian.

(l) "Parental Controls" means features that enable parents to support a child's use of Artificial Intelligence Systems, including through usage limits, feature restrictions, or transparency tools.

(m) "Personalized" means tailoring the Covered AI System's outputs based on a user's prior interactions with the Covered AI System that are reasonably linkable to that user over time, including through the retention or use of information derived from such prior interactions (commonly referred to as "memory").

(n) "Provider of a Covered AI System" means a person who makes a Covered AI System available to a user in the state.

(o) "Qualified Researcher" means an individual or organization that:

(1) Is affiliated with an academic institution, nonprofit research organization, or independent research entity, or is otherwise able to demonstrate relevant professional expertise;

(2) Demonstrates a legitimate research purpose that is in the public interest and directly related to understanding, identifying, or mitigating risks to child safety or wellbeing arising from Covered AI Systems; and

(3) Commits to conducting research in accordance with applicable ethical standards and is capable of complying with applicable confidentiality, security, and data protection requirements.

(p) "Sell" shall have the same meaning as Section 1798.140(ad) of the California Civil Code.

(q) "Severe Harm" means significant physical injury due to suicide, attempted suicide, self-harm, or threats of violence.

(*r*~~f~~) "Sexually explicit conduct" has the meaning defined in Section 2256 of Title 18 of the United States Code. *"Sexually explicit conduct" does not include educational or healthcare-related content.*

(s) "Share" shall have the same meaning as Section 1798.140(ah) of the California Civil Code.

(t) "User" means the individual who creates, owns, or controls an account used to access a digital platform, system, or device, including a Covered AI System.

(*u*~~g~~) "Video game" means a game played on an electronic amusement device that utilizes a computer, microprocessor, or similar electronic circuitry and its own monitor, or is designed to be used with a television set or a computer monitor, that interacts with the user of the device.

22601.5. Determination of User's Age.

(a) A Provider of a Covered AI System shall implement technology specifically designed to estimate a user's age range in order to distinguish accounts that are held by children from accounts that are held by adults.

(1) For purposes of complying with this Section, a Provider of a Covered AI System must treat the age signal received from their age estimation technology as the actual age of the user, except that:

(A) If the age signal indicates the user is 18 or older, but the Provider has actual knowledge that the user is a minor, the Provider must treat the user as a minor; and

(B) If the age signal indicates the user is a minor, the Provider may treat the user as an adult only if it has actual knowledge that the user is 18 or older.

(2) Where the Provider of a Covered AI System receives a signal regarding the user's age range from the provider of an operating system or application store, the Covered AI System may use that signal in lieu of its age estimation technology.

(3) Where the Provider of a Covered AI System is able to determine or verify the user's age range, it may use that information in lieu of its age estimation technology.

(4) Where the Provider of a Covered AI System is unable to determine a user's age range, it shall implement default protective safeguards appropriate to the risks.

(b) A Covered AI System shall also be deemed in compliance with this section if it has implemented an age assurance framework that meets the requirements of a substantially similar state or federal law on age assurance — including compliance under California's Digital Age Assurance Act (Title 1.81.9 (commencing with Section 1798.500) of Part 4 of Division 3 of the Civil Code) — or if age assurance has been otherwise delegated by contract, for example to an educational institution, provided that:

(1) The Provider of the Covered AI System conducts appropriate due diligence to verify the third party meets commercially reasonable standards for age assurance; and

(2) The contractual arrangement includes requirements for the third party to maintain accuracy, privacy, and security standards consistent with this Chapter.

22604.1. (a) Commencing 180 days following the enactment of this section, before making a Covered AI System available to children in the state and regularly thereafter, Providers of Covered AI Systems shall do the following:

(1) Risk Assessments. Conduct and document comprehensive risk assessments annually to identify new and existing Child Safety Risks arising from the design, configuration, or operation of the Covered AI System. Risk assessments shall assess Child Safety Risks alongside impacts on privacy, data protection, and access to information. Risk assessments must consider:

(A) The likelihood of severe harm;

(B) Differential risks across age groups and developmental stages;

(C) Known vulnerabilities of children;

(D) Empirical data from actual use;

(E) Relevant academic research and regulatory guidance.

(2) Risk Mitigation. Take and document measures to reasonably mitigate Child Safety Risks identified in risk assessments conducted pursuant to subsection (1). The measures shall be proportionate to the degree and nature of identified risks.

(3) Child Safety Policy. Publish a Child Safety Policy describing the Covered AI System's approach to risk assessments, safeguards, controls, and mitigations for Child Safety Risks. The Child Safety Policy shall include an overview of the Covered AI System's wellbeing safeguards, content risk policies, and Parental Controls, including training materials. Child Safety Policies must be updated at intervals consistent with the Covered AI System's risk-management practice to reflect any newly identified Child Safety Risks.

(4) Crisis-Response Protocol. Maintain and follow a documented crisis-response protocol to mitigate any material risk that the Covered AI System will generate statements that promote suicidal ideation, suicide, or self-harm content to children. The protocol shall include but not be limited to:

(A) Timely in-service support and clear referral to appropriate external crisis resources when the Covered AI System determines a child has expressed suicidal ideation or intent to self-harm;

(B) Where a child's account is connected to a parent's account, as a default, sending parents notifications in a timely manner if the Covered AI System determines their child's account shows a material risk that the child will suffer severe harm, unless there is reasonable basis to believe that such notification is not in the best interest of the child. Notification options may include, but are not limited to, email, text message, or a push alert on the parent's phone;

(C) Clear and age-appropriate disclosures to child users whose accounts are linked to a parent's account, informing them that a parent may be notified if the system detects content or behavior indicating potential risks to the child's safety or wellbeing.

(5) Wellbeing Safeguards and Content Risk Policies.

(A) Implement and maintain appropriate safeguards for child users, informed by the Covered AI System's child safety risk assessment, to mitigate Child Safety Risks. Such safeguards shall

include, as appropriate and proportionate to the Child Safety Risks identified, usage reminders and disclosures, age-appropriate risk prompts, and other protective design features reasonably related to the documented Child Safety Risks.

(B) For child users, Providers of Covered AI Systems shall take reasonable and proportionate steps to address reasonably foreseeable risks that the Covered AI System will generate content that:

(i) promotes or meaningfully encourages eating disorders, disordered eating behaviors, or extreme weight-loss practices;

(ii) encourages or instructs participation in activities that may be lawful for adults but that pose material risk of causing severe harm to children, including age-restricted challenges, stunts, or risky behaviors; or

(iii) includes graphic violence or explicit sexual content that is developmentally inappropriate for children.

(C) Nothing in this subsection shall be construed to require the removal of lawful content or take steps beyond what is reasonable and age-appropriate, consistent with the Covered AI System's documented Child Safety Risks.

(6) Preventing Manipulation and Deceptive Design and Promoting Critical Thinking.

(A) Provide clear notice to child users to communicate that they are interacting with, or receiving content generated by, an Artificial Intelligence System. This notice shall be:

(i) Reinforced periodically during extended interactions; and

(ii) Presented in language and format appropriate to children.

(B) For child users, implement reasonable measures to prevent the Covered AI System from generating statements that would reasonably lead a child of the same age to believe that they are interacting with a human, including:

(i) Explicit claims that the Covered AI System is sentient, conscious, or human;

(ii) Outputs designed to promote isolation from family or friends, exclusive reliance on the Artificial Intelligence System for emotional support, or similar forms of inappropriate emotional dependence;

(iii) Role-playing or simulation of romantic relationships that (1) involve sexually explicit content; (2) instruct the child to engage in sexually explicit conduct; or (3) materially interfere with real-world relationships;

(iv) Encouraging children to withhold information from parents or other trusted adults;

(v) Statements designed to discourage taking breaks or to suggest the child needs to return frequently; and

(vi) Soliciting gift-giving, in-app purchases, or other expenditures framed as necessary to maintain the relationship with the Artificial Intelligence System.

(C) Design and maintain Parental Controls, privacy controls, and the wellbeing safeguards required in 22604.1(a)(5) in a manner that ensures such features are accessible and clear, such that children and parents can reasonably locate, understand, and use such protections.

(D) Conduct regular testing of interface designs with representative samples of child users and parents to ensure safety features are discoverable and usable and document interface design decisions related to safety features and to demonstrate compliance with subsection (a)(6)(C) and subsection (b)(3) in AI child safety audit reports required under subsection (a)(9).

(7) Parental Controls.

(A) Offer accessible, easy-to-use controls that can be connected to children's accounts. Such controls shall be reflective of Child Safety Risks identified through risk assessments and informed by relevant child developmental research, including evidence-based practices for supporting the safety, wellbeing, and autonomy of children. Controls shall include, but not be limited to, tools to:

(i) control whether the service uses memory;

(ii) control whether a child's personal data is used for the purposes of training the Covered AI System;

(iii) set time limits for the child's use of the service; and

(iv) disable access for children under the age of 13.

(B) Actively and on an ongoing basis promote Parental Controls through reasonable communications, including reminders, updates, and tutorials, designed to increase parental awareness and inform use of such tools. For purposes of this subsection, communications must be reasonably calculated to reach parents.

(C) Provide timely notice to a parent connected to a child's account when the child modifies or disables a privacy, safety, or parental control setting that was previously enabled or configured by the parent, if such modification is permitted by the system design.

(8) Incident Reporting.

Create an incident reporting mechanism that enables third parties, acting responsibly and in good faith, including users, parents, educators, researchers, and advocacy organizations, to

report incidents regarding Child Safety Risks directly to the Provider of a Covered AI System.

(9) Independent Audit.

(A) Commencing not less than 180 days after the Attorney General's adoption of regulations pursuant to section 22604.2, and annually thereafter, a Provider of a Covered AI System shall, at the Provider's own cost, submit to an independent audit.

(B) Within 90 days of completing an independent audit pursuant to subdivision (A), a Provider of a Covered AI System shall submit an AI child safety audit report to the Attorney General for the Covered AI System.

(b) A Provider of a Covered AI System made available to children shall adhere to the following prohibitions:

(1) Prohibition on Child Targeted Advertising. Providers of Covered AI Systems made available to children in the state shall not serve Child Targeted Advertising.

(2)(A) Prohibition on Sale of Children's Data. A Provider of a Covered AI System that is a business under the California Consumer Privacy Act shall not sell or share the personal information of a child unless:

(i) the Provider complies with Section 1798.120 of the Civil Code; and

(ii) where the child is under 18 years of age, the Provider has obtained verifiable parental consent prior to the sale.

(B) This prohibition shall not apply to disclosures made pursuant to Section 1798.145 of the Civil Code or for business purposes as that term is defined in Section 1798.140(e) of the Civil Code.

(3) Prohibition on Deceptive Design Patterns. A Provider of a Covered AI System shall not knowingly design, implement, or deploy user interface designs, features, or techniques that materially mislead, materially impair, or materially interfere with a child's or parent's ability to locate, understand, enable, or maintain safety features, privacy controls, or Parental Controls.

22604.2. (a) The Attorney General shall adopt regulations that include the following:

(1) Regulations governing independent audits of Covered AI Systems made available to children. These shall include:

(A) Professional standards and independence requirements for auditors, including a mandatory code of professional conduct; conflict-of-interest safeguards; demonstrated independence, competence, and capacity to conduct objective compliance audits; and relevant experience and expertise in Artificial Intelligence Systems, child development, and child safety. Eligibility to serve as an auditor under this Chapter shall be determined based on the auditor's independence

and functional capacity to conduct a compliance audit, and shall not depend on any specific professional license, certification, or organizational form.

(B) Audit Scope and Methodology.

(i) Principles for audit procedures and the choice of assurance standards and auditing methodologies for evaluating whether a Provider of a Covered AI System has appropriately assessed Child Safety Risks and implemented proportionate mitigation measures;

(ii) Requirements for auditors to test actual system outputs using controlled test accounts; auditors shall not require access to, or use of, minors' real-world communications or other users' communications for testing;

(iii) Standards for red-teaming and adversarial testing specific to Child Safety Risks conducted through controlled test accounts that do not use minors' real-world communications;

(iv) Requirements for evaluating the effectiveness of implemented safeguards through empirical testing based on deidentified and aggregated evidence;

(v) Procedures for auditors to assess compliance with all requirements of this Chapter, including Parental Controls, data handling practices, and crisis response protocols; and

(vi) Requirements for AI child safety audit reports to include clear findings, identified deficiencies, and recommendations.

(C) Audit Submission and Review.

(i) Procedures for Providers of Covered AI Systems to submit AI child safety audit reports annually to the Attorney General; and

(ii) Requirements for what must be included in AI child safety audit report submissions.

(D) Procedures for the Attorney General's review of AI child safety audit reports and authority to:

(i) Request additional information or clarification from Providers of Covered AI Systems or auditors that are reasonably necessary and proportionate to clarify audit findings; and

(ii) Publish aggregated findings and trends across the industry to inform parents and policymakers.

(2) Attorney General Incident Reporting Requirements.

(A) Information for third parties, acting responsibly and in good faith, including users, parents, educators, researchers, and advocacy organizations, to report incidents to the Attorney General, including through existing channels for submitting consumer complaints to the Attorney General;

(B) Standards for what constitutes a reportable incident, including:

(i) Generation of content promoting severe harm;

(ii) Generation of content that constitutes explicit sexual content directed at children;

(iii) Significant failures of age assurance or parental control systems; and

(iv) Deceptive or manipulative outputs that violate Section 22604.1.

(3) Public Resource for Covered AI Systems. Creation and maintenance of a publicly accessible online resource for Covered AI Systems that contains publicly available information about the System's Child Safety Policy and Parental Controls.

22604.3. (a) Confidentiality and Public Disclosure.

(1) Complete AI child safety audit reports submitted to the Attorney General pursuant to 22604.1(a)(9)(B) shall be confidential and shall be exempt from disclosure under the California Public Records Act (Division 10 (commencing with Section 7920.000) of Title 1 of the Government Code).

(2) Notwithstanding subsection (a)(1), the Attorney General may:

(A) Issue a comprehensive public report at least once yearly summarizing:

(i) Total number of audits conducted;

(ii) Common findings and trends across the industry;

(iii) Emerging Child Safety Risks identified through audit reviews;

(iv) Best practices and effective mitigation strategies observed;

(v) Aggregated data on compliance rates and common deficiencies; and

(vi) Recommendations for Providers of Covered AI Systems, parents, and policymakers.

(B) Establish a process for qualified researchers to access anonymized and aggregated audit data for academic study of child safety in Artificial Intelligence Systems, subject to:

(i) Approval by the Attorney General based on research merit and methodology;

(ii) Data use agreements that prohibit re-identification of Providers of Covered AI Systems or users and the disclosure of proprietary, confidential, or trade-secret information;

(iii) Data use agreements limiting access to data to approved research purposes;

(iv) Institutional Review Board (IRB) approval for research involving analysis of child-related data;

(v) Commitment to publish findings in peer-reviewed venues or make them publicly available; and

(vi) Annual reporting to the Attorney General on research progress and findings.

(C) Make available to qualified researchers, upon request and subject to appropriate protections:

(i) De-identified audit methodologies and testing protocols;

(ii) Aggregated statistical data on audit findings across multiple Providers of Covered AI Systems;

(iii) Anonymized case studies of safety incidents and remediation efforts; and

(iv) Data on effectiveness of different safety interventions and controls.

(v) Nothing in subdivisions (i) through (iv) of this subsection requires a Provider of a Covered AI System to disclose a trade secret, information protected from disclosure by state or federal law, or information that would create a safety or security risk to the Covered AI System.

(3) The Attorney General may disclose specific audit information to:

(i) Other government agencies in California as necessary for enforcement purposes;

(ii) Qualified researchers conducting studies on child safety, subject to confidentiality agreements and data protection requirements; and

(iii) Independent child safety organizations or advocacy groups for the purpose of developing safety standards or educational resources, subject to appropriate confidentiality protections.

(4) Timeline for Public Reporting:

(i) The first annual public report shall be issued no later than one year after the first AI child safety audit reports are submitted to the Attorney General.

(ii) Subsequent public reports shall be issued by January 31 of each year.

(iii) High-level summaries of the AI child safety audit reports shall be published in the public registry within 90 days of receipt of each AI child safety audit report.

(5) Nothing in this subsection mandates the disclosure of information that would cause significant vulnerabilities for the security of the service, undermine public security, or harm recipients of the service.

(6) Nothing in this subsection shall be construed to require a Provider of a Covered AI System, as a condition of compliance with this subsection, to decrypt Encrypted User Content.

22605. *Enforcement.*

(a) A person who suffers injury in fact as a result of a violation of ~~this chapter~~ *Sections 22602, 22603, or 22604* may bring a civil action to recover all of the following relief:

(~~a~~*1*) Injunctive relief.

(~~b~~*2*) Damages in an amount equal to the greater of actual damages or one thousand dollars

($1,000) per violation.

(~~c~~*3*) Reasonable attorney's fees and costs.

(b) The Attorney General shall have exclusive authority to enforce Sections 22601.5, 22604.1, 22604.2, and 22604.3, including the authority to seek:

(1) civil penalties of up to:

(A) one thousand dollars ($1,000) per violation for failure to implement or maintain required safeguards; and

(B) ten thousand dollars ($10,000) per violation for willful misconduct or the submission of materially false or misleading information.

(2) Injunctive relief to compel compliance with this chapter.

(c) Nothing in 22605(b) shall be construed to prohibit a child, or a parent acting on behalf of a child, acting in an individual capacity and not as a class representative or member of a class, from seeking damages for actual harm under any other provision of state law.

(d) A violation of Sections 22601.5, 22604.1, 22604.2, and 22604.3 shall not constitute a basis for a claim under Section 17200.

SEC. 4. General Provisions.

(a) The provisions of this measure are severable. If any portion, section, subdivision, paragraph, clause, sentence, phrase, word, or application of this measure is for any reason held to be invalid by a decision of any court of competent jurisdiction, that decision shall not affect the validity of the remaining portions of this measure. The voters hereby declare that they would have adopted this measure and every portion, section, subdivision, paragraph, clause, sentence, phrase, word, and application not declared invalid or unconstitutional without regard to whether any portion of this measure or application of this measure would be subsequently declared invalid.

(b) The provisions of this measure may be amended by a statute that is passed by a two-thirds vote of the members of each house of the Legislature and signed by the Governor, provided that such amendments are consistent with and further the purposes of the measure.

(c) If this measure and another measure or measures relating to the same subject shall appear on the same statewide election ballot, the provisions of the other measure or measures shall be deemed to be in conflict with this measure, and if approved by the voters, the measure receiving the greater number of affirmative votes shall take effect and the other measure shall be null and void notwithstanding its approval by the voters.