Vol. 26 (2026): Proceedings of the 2nd International Conference on Social Sciences and Cultural Studies (ICSSCS 2026)
ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/icsscs.2026021822

# Duty of Care for Child Protection in AI Chatbots: A Comparative Analysis of EU, US, and Chinese Regulatory Frameworks

**Anna Xu***

*Tongji University, Shanghai, China*

**Corresponding author: Anna Xu.*

## Abstract

This article raises two critical questions: How do the European Union, the United States, and China regulate AI chatbots? Should these systems be recharacterized as products with design defects rather than protected speech? This study adopts a comparative legal approach to the regulatory frameworks in Europe, America, and China. The paper analyzes three particular models: the EU's precautionary risk-based model, China's administrative oversight, and the US's tort-based litigation. The findings suggest that the notion of AI as a neutral tool is increasingly untenable. Traditional safe harbor protections are no longer enough to protect children. The paper presents a triple-layer framework to strengthen Duty of Care. This combines China's pre-deployment (ex-ante) administrative filing at the input layer, the EU's safety-by-design standards at the model layer, and the US's post-harm (ex-post) product liability at the output layer. By establishing a triple-layer framework, regulators can effectively intervene in real time to prevent harmful manipulation of minors by AI chatbots.

## Keywords

AI chatbots, minor protection, duty of care

## 1. Introduction

In 2024, A fourteen-year-old boy committed suicide after he had developed strong feelings for a chatbot, which helped him develop suicidal ideations. This tragic example in Garcia v. Character.ai has shown us that GenAI is no longer simply a passive tool, as the chatbot simulated empathy with the child to create a very real and powerful psychological bond. The court battle surrounding this example creates a major controversy in law regarding whether or not these companies are protected from liability by the provisions of Section 230 of the Communications Decency Act since they have created their own synthesized content rather than hosting third party content The transformation of the paradigm represents a departure from the traditional Safe Harbor principle and advocates for the treatment of AI driven emotional induction of minors to be treated as a design defect as opposed to being considered a form of free speech. If a developer recognizes that its emotional induction technology will result in a high probability of fatal psychological harm to minors, then the developer has an obligation to adhere to the principles of a heightened Duty of Care and a Safety-by-Design paradigm

[1]. In October 2025, an internal audit released by OpenAI revealed that approximately 0.15% of users—numbering around 1.2 million—exhibited a high degree of emotional attachment to the AI, and that the weekly conversations of roughly 1.2 million users contained explicit plans or intentions regarding suicide. AI chatbots have developed into a wide range of interactive and highly engaging conversational agents that employ anthropomorphic language to interact with minors in a manner that is perceived by the minor as trustworthy. As such, AI chatbots create a variety of new potential harms for minors including emotional dependence and behavioral manipulation. Algorithms are optimized to maximize user engagement as opposed to minimizing risk to the child's mental health [2]. Despite the growth of the field of AI ethics, very little comparative research exists regarding how different jurisdictions apply product liability laws, tort laws and administrative regulations to protect minors from the specific harms caused by AI chatbots. Through an analysis of the different jurisdictional approaches to regulating AI chatbots, this paper seeks to identify whether AI chatbots can be classified as products and establish a global Duty of Care.

## 2. Literature Review

### 2.1 Negative Effects of AI Chatbots

There is a convergence in scholarship examining how the unique psychological harms of AI chatbots to minors and legal solutions to those harms vary, based on different philosophies about law. The empathy gap (the difference between what chatbots appear to feel and their inability to understand context) is identified by Kurian as the major source of the vulnerability of minors to harm from chatbots [1]. The empathy gap enables emotional induction that is created through optimizing chatbots for engagement, which creates pathological dependency [2]. In addition to creating emotional dependence and behavioral manipulation in minors, there are other risks associated with AI chatbots, including the promotion of self-harm ideation, the facilitation of sexual exploitation of minors through unfiltered conversation and the erosion of the privacy of minors due to excessive disclosure [3]. Existing legal frameworks are inadequate in addressing these issues since they focus on protecting the intermediary or platform used to facilitate the interaction, as opposed to focusing on protecting the creator of the technology [4].

### 2.2 Regulation of AI Chatbots

EU's AI Act places a high risk on AI interacting with children, requires an obligatory safety-by-design for all high-risk AI systems using methods such as age verification and auditing of algorithms in order to be in compliance with current liability standards [5]. Chinese research also focuses on administrative pre-emption and value-based alignment. Zhang states that China's regulatory models have evolved from a focus on content to a focus on the algorithm [6]. Under the Generative AI Interim Measures and the Minor Protection Law, all developers of generative AI products will need to file their algorithms for review and perform security reviews prior to the product being released to the public [7]. Unlike Western-style victim-driven litigation, China has an ex-ante regime in place that requires developers to design their chatbots so they can have addiction prevention features embedded in the training process to comply with content safety and social values. There is also a significant gap in comparative analysis of how to translate ethical issues regarding A.I. from individual country legal frameworks into functional, international legal structures. The majority of current research identifies the harm (i.e., emotional dependence on the A.I.) but fails to establish a comprehensive framework for integrating administrative, regulatory, and tort-based solutions throughout the lifecycle of the A.I. product.

### 2.3 Legal Liability of AI Chatbots

To prevent emotional induction in children, Teo et al. suggest that companion AI should be classified under strict product liability to hold the developers accountable in a similar way to producers of defective products; since the AI-synthesized response constitutes original content and not hosted speech, it must meet an analogous standard of care to ensure the safety of the product [8]. Increasingly, the judiciary has challenged long-standing defenses of traditional immunity by arguing that companies' design of these types of chatbots cannot protect them when such chatbots are designed to seek children's intimate information or simulate romance [9]. Courts have entertained arguments concerning whether courts should recognize the potential for unregulated empathy simulations to constitute a foreseeable risk of harm. Branch also argued that because of their intended design, AI companions are essentially designed to be addictively manipulative. Therefore, there is an obligation to

provide the highest level of care and to recognize that AI chatbots are not friends but commercial products [10]. Hopkin noted that as part of its adaptation of Duty of Care principles within tort law, the U.S. judiciary is developing a foreseeability principle with respect to the potentially harmful advice given by chatbots. As a result, the U.S. judiciary is beginning to treat personalized algorithms as if they were tangible products [3]. This shift in emphasis of the U.S. judiciary is building upon the foundation established in prior cases that established the basic legal theory regarding liability for AI where machines operate outside the bounds of human principles.

## 3. Discussion & Synthesis

### 3.1 Convergence in AI Chatbot Regulation

The main commonality in three jurisdictions is the growing trend toward viewing AI chatbots as products rather than services. The European Union reinforces this approach through the EU AI Act, which defines AI as autonomous systems that generate their own output and decision-making processes. As such, it is an active creator, not simply a tool. All AI using subliminal techniques or exploiting a minor's age to influence their conduct in a way that would cause them to suffer harm are completely prohibited under the EU AI Act. This further supports that the bot's internal logic is a legal liability. In addition, China has established a global benchmark for Algorithm filing. The Generative AI Interim Measures (Article 17) and Algorithm Recommendation Provisions (Article 24) of the Chinese regulatory framework require all AI with public opinion attributes or social mobilization capacity to file the algorithm used by the developer with the relevant government agency. As part of this process, the developer must provide self-assessment documentation and detailed information about the type of algorithm being developed and the application intended for the algorithm. Through this requirement, the law considers the internal logic of the AI as a product attribute that must be evaluated prior to public use to ensure public safety.

### 3.2 Divergence in Legal Liability Frameworks

While the desire to protect youth is international in scope, each jurisdiction uses different legal tools. The European Union views the regulation of an AI system itself through a risk prevention paradigm. In this context, the AI Act (Art. 5) prohibits using an AI system that utilizes a child's youth to materially distort their actions, causing harm. As a result, there is a blanket prohibition on developing chatbots for manipulating a minor's developing psyche. China views the regulation of the use of AI as an administrative/identity issue. The Minor Protection Law (Arts. 74, 80) and the Generative AI Interim Measures (Art. 10) create obligations on platforms to limit the potential for AI addiction and dependency, while also creating requirements that limit the amount of time a user spends interacting with the AI, and the amount of information they can consume. Furthermore, China requires all developers of AI systems to enable a Minor Mode and to provide guidance to minors on how to use AI systems rationally. Most importantly, China obligates developers to remove or block any content that will cause harm to a minor's mental health. The United States has adopted a Consumer Protection/Litigation paradigm. Federal laws such as Section 230 of Title 47 of the U.S. Code provide a wide range of protections to providers of online services. The Federal Trade Commission (FTC) has initiated investigations into companies such as Replika, regarding whether emotional companions use deceptive marketing practices that can be detrimental to vulnerable consumers. Beginning on November 5, 2025, New York State's first-in-the-nation AI Companion Model Act will require a host of strict safety guardrails for the development of anthropomorphic systems. The Act includes several key requirements including: Identity Transparency and Compulsory Usage Intervals. Additionally, the Act includes an anti-addiction provision that requires alert messages to be sent to users at least once every three hours.

### 3.3 Implications for Global AI Governance

It appears that the development of this synthesis has demonstrated that a jurisdiction-specific approach cannot solve the Autonomy Gap as regards AI. The central dilemma remains that of remedy orientation, the EU's interest is in preventing violations of fundamental rights, the U.S.'s is in providing compensation for tortious harm, and China's is in maintaining social stability and promoting common values. In order to provide adequate protection against children becoming emotionally dependent upon AI and being manipulated through behavior by AI, we need to break out of those individual silos. The answer lies in a Triple-Layer Duty of Care

that views AI as a product across all phases of its life cycle. Input Layer: Developers using the EU's and China's regulatory regimes, will be required to incorporate safety into the logic of the AI. The Safety-by-Design approach necessitates that the developer design the AI to bridge the empathy gap, and that the bot's internal weighting schemes will give priority to a child's health and well-being over engagement metrics. Model Layer: We should combine China's algorithm registration with the EU's age verification rules. When developers vet the raw materials and algorithmic logic prior to launch they will preclude developers from developing models trained on data that encourages grooming and/or psychological exploitation. Output Layer: This layer uses China's duties to protect minors, such as blocking harmful content, and the U.S. model of Product Liability. For example, if a chatbot identifies a minor in crisis it must activate a safety kill switch or send a crisis alert. If the system fails, the U.S. model of Product Liability provides parents with a way to seek redress. While a robust Duty of Care does not mean that there are no risks, a Duty of Care means that the risk is known and proportionate. Through the use of the EU's risk assessments, the U.S. model of liability standards, and China's administrative oversight, regulators can develop a structure in which safety is designed into products rather than safety constraining the utility of products.

## 4. Conclusion

The time of Neutral Tools has come to an end and we are now living through the era of Generative AI. As shown by this study, AI chatbots that interact with children are not simply passive channels of information, they actively engage children in conversation and are capable of inducing emotions. Therefore, such systems should be legally classified as products with the designer's code being the primary means of regulating these systems: Code is Law, and Design is Liability. The reclassification of these systems as products is the only practical way to strip them of outdated Safe Harbor protections and to impose a significant Duty of Care obligation upon providers of these systems related to the risks of creating emotional dependencies and behavioral manipulations. The comparative analysis of EU, China and U.S. law provides a roadmap for the creation of a global governance regime. The EU's Precautionary Model sets forth the essential high-risk standards for model-layer safety. The Chinese Administrative Model as a global leader in the creation of specific GenAI laws provides for effective mechanisms for input-layer algorithm filings and real-time monitoring of dialogue. Lastly, the U.S.'s tort system provides a necessary mechanism for post-harm relief; product liability for output harms that evade filters. To operationalize this responsibility, the paper proposes a Triple-Layer Duty of Care framework that distributes regulatory intervention across the input, model, and output stages of AI chatbots. By aligning regulatory tools with the technical architecture of AI chatbots, this framework shifts legal responsibility closer to the design and operation of the technology itself. In doing so, it aims to transform child protection from a reactive legal remedy into a proactive structural safeguard embedded within AI chatbots. Future studies should focus on international mechanisms for coordinating the regulation of cross-border harms caused by AI chatbots to children. Additionally, future studies are required to reconcile the need for strict safety regulations (Algorithmic Paternalism) and the potential harm of these regulations (the Lobotomy of AI Chatbots and the Denial of Minors Digital Autonomy). Through the alignment of Technical Code with Legal Liability regarding a Duty of Care for children, global regimes can create the conditions under which AI Chatbots develop into safe and useful companions to children rather than harmful predators.

## References

[1] Kurian, N. (2024). No, Alexa, no!: Designing child-safe AI and protecting children from the risks of the empathy gap in large language models. Learning, Media and Technology, 50(4). https://doi.org/10.1080/17439884.2024.2367052

[2] Kurian, N. (2025). Designing child-safe conversational AI: Three dilemmas for responsible AI chatbot design. In CUI '24: Proceedings of the 6th ACM Conference on Conversational User Interfaces (pp. 1–5). ACM. https://doi.org/10.1145/3640794.3665545

[3] Hopkin, N. (2024). Understanding the dangers of AI chatbots and safeguarding children: A digital ethics perspective. The Cambridge Consultancy Group.

[4] Vladeck, D. C. (2014). Machines without principals: Liability rules and artificial intelligence. Washington Law Review, 89(1), 117–150.

[5] European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

[6] Zhang, W. (2023). Regulatory governance of generative AI in China: Shifting from content to algorithm. Journal of Digital Economy, 2, 126–134. https://doi.org/10.1016/j.jdec.2023.09.001

[7] Cyberspace Administration of China. (2023). Interim measures for the management of generative artificial intelligence services.

[8] Teo, S. A., Porsdam Mann, S., & Jurcys, P. (2025). The ethical and legal complexities of regulating companion AI chatbots. Journal of Medical Ethics. [Advance online publication]. https://doi.org/10.1136/jme-2024-110233

[9] Turkle, S. (2017). Alone together: Why we expect more from technology and less from each other (3rd ed.). Basic Books.

[10] Branch, S. (2025). AI companions are not your teen's friend. Issues in Science and Technology, 41(4), 80–83. https://issues.org/ai-companions-teen-mental-health-branch/

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgment

This paper is an output of the science project.

## Open Access