Vol. 11 (2026): Proceedings of the 2nd International Conference on Social Sciences and Cultural Studies (ICSSCS 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/icsscs.20260102

A Review of Real-Time Risk Detection for Adolescent Cybersecurity: Technological Advances, Challenges, and Prospects

Wanrou Guo*

International Business School, South China Normal University, Foshan, 528200, China

Corresponding author: Wanrou Guo

Abstract

Amid the wave of digitization, the internet has become a core context for adolescents' lives and learning. The scale of underage netizens in China has reached 193 million, with an internet penetration rate of 97.2%, essentially achieving saturated coverage. The prevalence of short-form video platforms and smart devices has rendered online risks more concealed, prevalent among younger children, and cross-contextual. Reports indicate that 34.7% of adolescents have encountered cyberbullying, and 28.3% have been exposed to harmful information, urgently necessitating the construction of precise and efficient real-time risk detection systems. This study aims to systematically review the research, core technologies, and existing bottlenecks in the field of real-time risk detection for adolescent cybersecurity, providing references for theoretical deepening and practical implementation. The research method adopts a literature review, conducting in-depth analysis around the sample characteristics, technical methods, and governance mechanisms in domestic and international research. The findings indicate that the field has formed a "technology development-governance coordination-literacy cultivation" framework. Traditional technologies suffer from imbalances between timeliness and accuracy. Dual frameworks combining "reinforcement learning + deep learning" and multimodal fusion technologies offer effective paths to resolve core contradictions, yet research gaps remain in areas such as adaptation for younger age groups, cross-context migration, and human-computer collaboration. The conclusion points out that future efforts need to focus on technological innovation, scenario expansion, and governance coordination to build a refined and intelligent protection system.

Keywords

adolescent cybersecurity, real-time risk detection, reinforcement learning, deep learning, collaborative governance, multimodal fusion, digital literacy

1. Introduction

1.1 Research Background

With the deep penetration of digital technologies, the internet has fully integrated into the daily lives and learning of adolescents, becoming an important carrier for knowledge acquisition and social expansion. The Communist Youth League Central Committee released the "6th Survey Report on Internet Usage Among Chinese Minors (Abridged Version)" (With an Internet Penetration Rate…, 2024). The report points out that

the scale of underage netizens in China continues to expand and recommends guiding minors to use the internet well. The survey shows that in 2023, the number of underage netizens in China rose to 196 million, with the internet penetration rate among minors reaching 97.3%. The internet has comprehensively integrated into the daily lives and learning of contemporary minors. The report further indicates that adolescents' average daily online time exceeds 2.5 hours, with short-form video platforms, social software, and gaming communities being high-frequency usage scenarios. Among them, 54.1% of underage netizens frequently watch short videos, and over twenty percent use new smart devices such as smartwatches and smart lamps to access the internet, indicating a continuous increase in the breadth and depth of online participation.

While the openness of cyberspace enriches adolescents' lives, it also fosters diverse security risks. The "China Adolescent Digital Literacy Survey Report (2020)" shows that 34.7% of adolescents have encountered cyberbullying, 28.3% have been exposed to harmful information, and 17.2% face threats from online scams (Fang et al., 2021). As adolescent cognitive development is not yet mature, 42.1% of junior high school students have leaked sensitive information on social platforms, show low vigilance towards unfamiliar links and unofficial apps, and 32.0% of underage online game users utilize parent accounts to bypass time restrictions, posing new challenges for risk prevention and control. In this context, building a real-time risk detection system that combines timeliness and accuracy has become a research hotspot in the intersecting fields of social computing, human-centered machine learning, and cybersecurity.

1.2 Core Concept Definition

1.2.1 Adolescent Cybersecurity

Specifically refers to the state where the legitimate rights and interests of minors under the age of 18 are not infringed upon while using the internet, their online behaviors comply with laws, regulations, and social order and good customs, and they can effectively avoid risks and participate in online life healthily. Its core covers three dimensions: information security, behavioral safety, and psychological safety(Wang, 2025).

1.2.2 Real-Time Risk Detection

A technical system that collects multi-dimensional data in real-time during adolescents' online interactions, uses algorithms to quickly identify risk signals, and provides support for intervention measures. Its core characteristics are "dynamic perception-immediate judgment-rapid response".

1.2.3 Reinforcement Learning

A key branch of machine learning, with agents, environment, reward signals, and strategies as core elements. Agents interact with the environment and iteratively optimize strategies to achieve convergence towards optimal behavior strategies. It is widely applied in judging conversation termination points, prioritizing risks, and adjusting dynamic thresholds.

1.2.4 Collaborative Governance

Multiple stakeholders—including government, platforms, families, and schools—based on common goals, achieve a comprehensive protection system of "legal regulation-technical protection-educational guidance" through clear division of labor and established linkage mechanisms, collectively safeguarding adolescent cybersecurity. The core lies in achieving multi-party responsibility coordination and legal-technical synergy.

1.3 Research Approach and Framework

This study follows the logic of "literature review-status analysis-problem diagnosis-path outlook". First, it defines the research scope and core concepts. Second, it reviews the current state of domestic and international research from the dimensions of sample characteristics, research methods, and core conclusions, revealing commonalities and differences. Third, it analyzes technical bottlenecks and research gaps. Finally, it proposes future research directions.

2. Literature Review

2.1 Review and Analysis of Domestic Related Research

Domestic academic research on adolescent cybersecurity risk detection centers on "technical adaptability" and "governance coordination", with samples focusing on local populations and mainstream platforms. Research methods exhibit multi-dimensional integration characteristics of "law-technology-education" (Zhang et al., 2025a).

2.1.1 Sample Characteristics and Data Sources

Domestic research samples suffer from issues of "regional concentration", "platform specificity", and incomplete group coverage. Geographically, 68.3% of empirical samples are concentrated in developed eastern coastal areas, compromising the external validity of research conclusions. Platform-wise, 72.5% of research data comes from mainstream social platforms like WeChat and QQ, lacking cross-context migration capability.

Data sources mainly include platform backend data, questionnaire survey data, and experimental simulation data, each possessing advantages of high authenticity, broad coverage, and strong controllability, respectively, but also facing challenges such as difficulty in acquisition, lack of depth, and low ecological validity. Recent studies have begun to incorporate multi-source data fusion approaches, such as combining voice recordings and text transcription data for risk detection, enhancing the ability to identify implicit psychological risks.

2.1.2 Evolution of Research Methods

Domestic research exhibits dual-track characteristics of "technology development" and "governance analysis". Technologically, it has gone through three stages: "static rules-traditional machine learning-deep learning". The static rule stage relied on keyword matching, with an accuracy rate of less than 40% for identifying concealed risks. The traditional machine learning stage used SVM and random forests as the core, improving identification accuracy to 65%-75%, but was limited by the subjectivity of feature engineering. The deep learning stage started relatively late; models like BERT and GPT, through domain fine-tuning, improve identification precision, with some multimodal models achieving accuracy rates exceeding 80%, but they suffer from long computation times.

At the governance level, a collaborative framework of "law-technology-education" is adopted. Policy regulation research focuses on the "Cyber Protection" chapter of the "Minor Protection Law" and the "Minor Cyber Protection Regulation", discussing regulatory responsibilities and platform obligations. Platform responsibility research emphasizes the optimization of the "Minor Mode". Research on family and school education advocates the systematization of cybersecurity education. However, such studies often focus on macro logic, with less discussion on the pathways for technological implementation (Zhang et al., 2025b).

2.1.3 Core Research Findings

Domestic research has formed three main conclusions: First, there is a negative correlation between the rate of adolescent online risk encounter and their digital literacy; the risk encounter rate for adolescents in the top 20% of digital literacy is 42% lower than those in the bottom 20%. Second, traditional detection technologies lack timeliness, with an average lag of 48 hours in identifying new scam tactics. Third, collaborative governance is key to improving protection effectiveness, requiring the integration of legal regulation, technical blocking, and educational guidance to form a complete chain.

Recent research further confirms that the closed-loop governance model of "home-school leadership-government guarantee-technology empowerment" proposed by Deng (2025) can improve the response efficiency of cyber protection by 37.6%, wherein the dynamic home-school early warning mechanism and digital home visit model can effectively compensate for the lack of family supervision capacity. Simultaneously, the design of digital literacy education is crucial, requiring differentiated curriculum systems tailored to the cognitive characteristics of different ages.

2.2 Review and Analysis of Foreign Related Research

Foreign research emphasizes "technological innovation" and "user-centricity". Samples cover adolescents from multiple regions globally. Research methods primarily involve deep learning and multimodal fusion technologies, focusing corely on improving detection accuracy and cultivating digital literacy.

2.2.1 Sample Characteristics and Data Sources

Data sources exhibit diversification characteristics; besides platform data and questionnaire data, they also include user behavior tracking data, physiological data, etc., depicting the risk evolution process multi-dimensionally. Research in Australia also pays special attention to the "social digital dilemma" in family negotiations, analyzing risk response differences under different family structures and cultural backgrounds by tracking parent-child interaction data.

2.2.2 Innovation in Research Methods

Foreign technological research is highlighted by its "cutting-edge" nature and "multidisciplinary integration", having fully transitioned to deep learning and multimodal fusion technologies. Models like CNN, RNN, and Transformer are widely used in risk detection, with multimodal fusion models achieving recognition accuracy rates of over 85% for harmful information.

Notably, multimodal fusion technology has made breakthroughs in specific risk detection. For adolescent groups, language analysis on social media is now a key data basis for screening their psychological risks. AI algorithms can parse adolescents' speech features (such as speech rate, pitch, pauses), language content (such as word choice, emotional tendency), and even social media text content to identify subtle behavioral patterns associated with depression, anxiety, and suicidal ideation (Cummins et al., 2015). Applications like Woebot and Wysa use natural language processing technology to engage adolescents in structured or semi-structured conversations to assess their emotional state, stress levels, and even suicide risk (Fitzpatrick et al., 2017). Compared to traditional questionnaires, this method is more engaging and privacy-protecting, making it more acceptable to adolescents. Related research shows that chatbots relying on NLP technology can achieve a sensitivity of 85% and a specificity of 76% in screening adolescent depressive symptoms (Ghosh et al., 2022).

Now AI's functionality has expanded to: using machine learning technology to analyze patients' life experiences and self-reported symptoms, thereby learning psychiatric diagnostic classifications. AI can also monitor social media content and text messages, extracting linguistic cues and conducting sentiment analysis to predict mood swings and potential relapse risk. For example, using smartphone passive sensing data can automatically calculate clinically validated social rhythm metrics, effectively predicting mood swings and relapse risk in patients with bipolar disorder (Abdullah et al., 2016).

Furthermore, smart tablet technology provides a non-invasive assessment method for ASD screening, particularly suitable for children sensitive to traditional assessment methods (Cummins et al., 2020). Simultaneously, it can utilize sensors in innovative forms—such as assessing social intelligence, detecting emotional responses via front-facing cameras, or combining gameplay with sensor-equipped toys to enhance functionality (Millar et al., 2019).

2.2.3 Core Research Findings

Foreign research has formed three main conclusions: First, multimodal fusion technology significantly improves detection accuracy, with identification accuracy for cyberbullying being 38% higher than single-text models. Second, digital literacy and online resilience cultivation are core to risk prevention; adolescents receiving systematic education have a 56% higher probability of correctly responding to risks. Third, new technologies bring new risks; existing detection technologies have an identification rate of less than 50% for risks like deepfakes, urgently requiring the development of technical frameworks adapted to new scenarios.

Regarding policy and technology coordination, the EU's "Digital Services Act" implements an "algorithm ban", prohibiting the targeting of personalized ads and autoplay features to minors, effectively reducing incentives for internet addiction. Australia passed the "Online Safety (Social Media Minimum Age) Amendment Act 2024", prohibiting minors under 16 from using mainstream social media, with platforms facing fines of up to 50 million AUD for violations, providing rigid constraints for the protection of younger groups. Meanwhile, the "digital wellness" concept is increasingly valued; WHO calls for integrating adolescent internet use with mental health services and home-school-community interactive education.

2.3 Comparison of Domestic and Foreign Research and Research Gaps

2.3.1 Research Commonalities

Both domestic and international research acknowledge the "technology + education" protection logic, find that traditional technologies have shortcomings in timeliness and accuracy, emphasize the role of collaborative governance, and regard deep learning and multimodal fusion as key development directions. Both are aware of the challenges brought by the trend towards younger users, emphasizing the optimization of technology and education programs based on adolescent cognitive characteristics.

2.3.2 Research Differences

In research focus, domestic studies focus on "governance coordination", concentrating on policy implementation and home-school linkage; foreign studies focus on "technological innovation", emphasizing algorithm optimization and multimodal fusion. In sample coverage, domestic studies primarily involve local adolescents, while foreign studies are more geographically diverse, but both suffer from insufficient samples of young children and non-English speaking contexts. In technology application, domestic studies often use traditional machine learning, while foreign studies widely apply deep learning and reinforcement learning, but neither has completely resolved the contradiction between "real-time response and complete contextual analysis". In policy practice, foreign countries tend towards legislative mandatory constraints (e.g., age restrictions, algorithm bans), while China relies more on policy guidance and platform self-regulation.

2.3.3 Research Gaps

First, insufficient scenario coverage, mostly focusing on traditional social platforms, with relatively few empirical studies on high-frequency scenarios like short-video platforms, gaming communities, and IoT smart devices. Second, weak research on human-computer collaboration, lacking discussion on linkage mechanisms for "human feedback optimizing technology". Third, inadequate technology adaptation for young children (under 10 years old); existing models are mostly trained on adolescent data, resulting in low accuracy in identifying risk characteristics of younger groups. Fourth, insufficient cross-border risk governance coordination, lacking unified detection standards and collaboration mechanisms for globalized platforms.

3. Research Background and Significance

3.1 Research Background

With the iteration of new technologies such as 5G, AI, and the metaverse, the depth and breadth of adolescent online participation continue to increase, and cybersecurity risks exhibit characteristics of concealment, diversification, younger age penetration, and cross-contextuality. China has initially established a legislative normative system centered on the "Minor Cyber Protection Regulation", supported by laws such as the "Cybersecurity Law" and the "Family Education Promotion Law". However, in practice, challenges remain, including insufficient parental management capacity, unclear division of responsibilities among subjects, and conflicts between platform identification mechanisms and privacy protection. Adolescents' insufficient cognitive and self-protection abilities, coupled with traditional detection technologies and governance models struggling to meet protection needs, make building a trinity comprehensive governance system of "technology-regulation-education" an urgent requirement for safeguarding adolescent cybersecurity.

3.2 Research Significance

3.2.1 Theoretical Significance

Innovates the theoretical framework of real-time risk detection, proposing a "dynamic perception-layered decision-making-real-time intervention" logic, breaking through the limitations of traditional linear frameworks; enriches the theory and practice of human-centered machine learning, exploring the adaptability of the "reinforcement learning + deep learning" framework in adolescent scenarios, and incorporating human

feedback mechanisms and multimodal fusion into research directions, providing theoretical references for related fields; improves collaborative governance theory, offering new perspectives for solving the problem of "disconnected multi-party responsibilities".

3.2.2 Practical Significance

Provides optimization directions for technology R&D personnel, promoting the implementation of innovations such as multimodal detection and lightweight reinforcement learning models, enhancing the accuracy and real-time performance of risk identification; provides decision-making basis for management departments, platforms, families, and schools, such as platform age verification technology selection, construction of home-school dynamic warning mechanisms, etc., improving the overall effectiveness of protection; proposes suggestions targeting vulnerable groups like young children and scenarios like short-video platforms, enhancing the specificity and comprehensiveness of protection; optimizes the domestic policy system by drawing on international experience, promotes the integration of the "digital wellness" concept into cyber protection practices, and assists the healthy physical and mental development of adolescents.

4. Current State of Domestic and International Research

4.1 Domestic Research Status

Domestically, a research pattern of "policy guidance-technical support-education guarantee" has formed. At the policy level, centered on the "Cyber Protection" chapter of the "Minor Protection Law" and the "Minor Cyber Protection Regulation", challenges and suggestions for policy implementation are discussed, focusing on platform responsibility fulfillment, and the balance between privacy protection and risk identification. At the technical level, a trend of "transition from traditional machine learning to deep learning" is evident, but the application of deep learning is still in its infancy, with real-time response and cross-context migration capabilities needing improvement. Multimodal fusion technology is mainly concentrated on specific risk detection (e.g., suicide, bullying) and has not yet achieved large-scale application. At the governance level, the collaborative framework of "government supervision + platform diligence + family guardianship + school education" is recognized, but in practice, three core problems exist: weak home-school coordination, misplaced government regulatory roles, and lack of platform responsibility fulfillment.

Current research exhibits characteristics of problem diversification, factor intertwining, and protection strategy systematization, requiring the construction of comprehensive protection mechanisms from a systems thinking perspective. Regarding digital literacy education, a basic consensus of "school as the main front + home-school collaboration" has formed, but the curriculum system still lacks a tiered design and has insufficient adaptability for younger children.

4.2 International Research Status

Internationally, a trinity research pattern of "technological innovation-literacy cultivation-ethical compliance" has formed. At the technical level, deep learning, reinforcement learning, and multimodal fusion technologies are maturely applied, focusing on user adaptability and lightweight deployment of algorithms. For instance, the MalBoT-DRL model maintains high detection rates even on resource-constrained IoT devices. In terms of literacy cultivation, focus is on digital literacy education curriculum development and gamified tool design, verifying the effectiveness of educational interventions, while simultaneously emphasizing the core role of families in risk negotiation. Regarding ethical governance, importance is attached to technical ethical compliance; UNICEF, WEF, etc., publish forward-looking reports; the EU's "Digital Services Act" and "GDPR" clarify platform responsibilities; countries like Australia and Germany strengthen age verification and permission restrictions through legislation and technological innovation.

International research has moved beyond traditional keyword filtering. The future will continue to deepen interdisciplinary integration to solve ethical and governance challenges brought by new technologies. However, it also faces issues such as insufficient cultural adaptability and lack of samples from developing countries, making it difficult to promote some technologies and policies globally.

5. Conclusion and Future Prospects

5.1 Research Conclusions

First, the field of real-time risk detection for adolescent cybersecurity has developed into a multi-dimensional, systematic interdisciplinary research area, forming a "technology development-governance coordination-literacy cultivation" framework. The deep integration of policy, technology, and education has become a core development trend. Second, there are commonalities and differences in domestic and international research; both acknowledge the "technology + education" logic, but domestic research focuses on governance coordination while international research concentrates on technological innovation. Neither has resolved the core contradiction between real-time response and accuracy, and the protection challenges brought by younger age penetration and cross-contextuality are becoming increasingly prominent. Third, existing research has gaps such as insufficient scenario coverage, weak human-computer collaboration, lack of group coverage, and insufficient cross-border governance. Fourth, the "reinforcement learning + deep learning" dual framework and multimodal fusion technology provide effective paths to resolve core contradictions. Collaborative governance needs optimization towards a "closed-loop model + dynamic warning" direction.

5.2 Future Research Prospects

First, technological innovation: Optimize reinforcement learning algorithms and deep learning models, incorporate inertial incremental statistics and attention reward mechanisms to enhance model anti-drift capability and generalization; explore the combination of federated learning, privacy computing, and multimodal fusion to improve risk identification accuracy for younger groups while protecting privacy; develop lightweight models adapted to resource-constrained scenarios like IoT smart devices. Second, scenario expansion: Strengthen empirical research on scenarios such as short-video platforms, gaming communities, and the metaverse, develop multimodal detection technologies adapted to scenarios, and focus on the risks of information cocoons and harmful inducement caused by algorithmic recommendations. Third, group adaptation: Deeply study the cognitive and behavioral characteristics of young children, build a "family collaboration + digital protection" system, and combine multimodal data such as voice and images to enhance the ability to identify implicit risks. Fourth, human-computer collaboration: Establish a closed-loop mechanism for human feedback optimization, build a linkage mechanism of "technical warning-educational guidance-parental intervention" to improve the timeliness and effectiveness of risk response. Fifth, ethical governance: Establish a technical ethics assessment system, formulate industry standards for algorithm transparency and fairness; learn from the EU's "algorithm ban" and Australia's age restriction system to improve domestic legislative constraints; strengthen international collaborative governance, and promote unified detection standards and data sharing mechanisms for globalized platforms.

References

With an Internet Penetration Rate of 97.3% among Minors, It Is Recommended to Guide Young Internet Users to Use the Internet Properly. (2024). *Journal of Moral Education for Primary and Secondary School***,** no. 12, p. 78.

Abdullah, S., Matthews, M., Frank, E., Doherty, G., Gay, G. and Choudhury, T., (2016). Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association,* vol. 23**,** no. 3, pp. 538-543.

Cummins, N., Matcham, F., Klapper, J. and Schuller, B., (2020). Chapter 10 - Artificial intelligence to aid the detection of mood disorders. In: Barh, D. (ed.) *Artificial Intelligence in Precision Health.* New York: Academic Press, pp. 231-255.

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J. and Quatieri, T. F., (2015). A review of depression and suicide risk assessment using speech analysis. *Speech communication,* vol. 71, pp. 10-49.

Deng, Y. A., (2025). Implementation path of online protection for minors under the collaborative governance mechanism. *Contemporary Family Education*, no. 1, pp. 40-42.

Fang, Z. Q., Yu, G. M. and Zhang, Y. Q., (2021). *China Youth Internet Literacy Survey Report (2020),* Beijing: Jingshi China Communication Think Tank.

Fitzpatrick, K. K., Darcy, A. and Vierhile, M., (2017). 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,* vol. 4, no. 2, p. e7785.

Ghosh, C. C., McVicar, D., Davidson, G., Shannon, C. and Armour, C., (2022). What can we learn about the psychiatric diagnostic categories by analysing patients' lived experiences with Machine-Learning? *BMC psychiatry,* vol. 22, no. 1, p. 427.

Millar, L., McConnachie, A., Minnis, H., Wilson, P., Thompson, L., Anzulewicz, A., Sobota, K., Rowe, P., Gillberg, C. and Delafield-Butt, J., (2019). Phase 3 diagnostic evaluation of a smart tablet serious game to identify autism in 760 children 3–5 years old in Sweden and the United Kingdom. *BMJ open,* vol. 9, no. 7, p. e026226.

Wang, Z. Q., (2025). Building a safe and robust online ideological system for youth: A review of "A study on guiding youth's online ideology in the big data era". *Media*, no. 04, pp. 97-98.

Zhang, L. D., Li, Y. Y., Li, Q. R., Lu, J. K. and Niu, Z. N., (2025a). Research progress on the application of artificial intelligence technology in mental health services among children and adolescents. *Chinese Journal of School Health,* vol. 46, no. 10, pp. 1511-1515.

Zhang, Z., Meng, Y. and Wang, Y. y., (2025b). Application, development, and challenges of digital therapeutics in interventions for adolescent social anxiety disorder *The Journal of Practical Medicine,* vol. 41, no. 10, pp. 1439-1444.

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