

Ethical Risks and Mitigation Mechanisms of Artificial Intelligence in College Students' Psychological Crisis Intervention

Gujing Li^{1*}

¹ Zhejiang Dongfang Polytechnic, Wenzhou, China

Email Address

241977969@qq.com (Gujing Li)

*Correspondence: 241977969@qq.com

Received: 29 January 2026; **Accepted:** 2 March 2026; **Published:** 14 April 2026

Abstract:

With the deep integration of artificial intelligence (AI) into college students' psychological crisis intervention (PCI), AI has significantly improved the accuracy of risk identification and intervention efficiency, but also triggered multiple ethical risks such as privacy leakage, algorithmic bias, ambiguous responsibility, and subjectivity alienation (referring to the weakening of students' subjective initiative and humanistic care caused by over-reliance on AI technology). Based on data ethics, algorithm ethics, educational ethics, and risk governance theory, this study identifies ethical risks from technical, institutional, and cognitive dimensions, and constructs a 'technology-institution-ethics-practice' four-in-one mitigation mechanism. This research enriches the ethical research dimension of the interdisciplinary field of 'AI + psychological education' and provides theoretical support and practical paths for the standardized and sustainable development of AI-enabled PCI in colleges worldwide, realizing the dynamic balance between technological empowerment and ethical constraints.

Keywords:

Artificial Intelligence (AI), Psychological Crisis Intervention (PCI), Ethical Risks, Mitigation Mechanisms, College Mental Health

1. Introduction

Policy documents such as the New Generation Artificial Intelligence Development Plan and the Guidelines for Mental Health Education of College Students have explicitly promoted the in-depth integration of AI with education and mental health fields, laying a solid policy foundation for technological applications. Meanwhile, Generation Z college students exhibit psychological crisis characteristics of "high concealment, high suddenness, and high public opinion sensitivity". Traditional intervention models struggle to meet the demands for precise identification and personalized response, making AI a crucial solution to this predicament. However, its application involves key links such as multi-source data collection, algorithmic decision-making, and intelligent intervention, leading to the increasing prominence of

ethical issues including privacy protection, algorithmic fairness, and accountability definition. Consequently, a systematic risk mitigation mechanism is urgently required.

This research not only enriches the ethical research dimension of the interdisciplinary field of “AI + psychological education”, improves the ethical theoretical system of technology-empowered education, and provides a novel analytical framework of “data-model-scenario-mechanism” for interdisciplinary ethical research. Additionally, it accurately identifies core ethical risk points and constructs an operable “four-in-one” mitigation mechanism, which safeguards the legitimate rights and interests of college students, enhances the standardization and credibility of PCI in higher education institutions, and promotes the sound development of technology in the mental health domain.

2. Definition of Core Concepts and Theoretical Foundations

This section defines four core concepts related to AI application in college students' PCI, and relies on data ethics, algorithm ethics, educational ethics, and risk governance theory to lay a solid theoretical foundation for subsequent research, providing conceptual support and academic guidance.

2.1. Definition of Core Concepts

2.1.1. Artificial Intelligence Technology (in the context of PCI)

Refers to a technical system in the weak AI stage, encompassing machine learning, natural language processing, big data and cloud computing, computer vision, and AR/VR technologies, applied throughout the entire process of psychological crisis identification, assessment, and intervention. Its core function is to assist professionals in improving the precision and efficiency of interventions.

2.1.2. Psychological Crisis Intervention

Refers to systematic work implemented by colleges and universities for students experiencing psychological imbalance. It aims to help them restore psychological equilibrium and prevent crisis escalation through a series of professional activities, such as early warning screening, emotional counseling, and professional referral.

2.1.3. Ethical Risks

Refers to the potential likelihood that the application of AI technology in PCI may infringe upon students' legitimate rights and interests, violate educational ethical norms, and harm public interests due to inherent technical characteristics, inadequate institutional norms, or improper operation and execution.

2.1.4. Mitigation Mechanisms

Refers to a systematic safeguard system established to prevent and resolve ethical risks, covering technical standards, institutional norms, ethical guidelines, and practical processes. Its core goal is to achieve a dynamic balance between technological application and ethical requirements.

2.2. Theoretical Foundations

2.2.1. Data Ethics Theory

Emphasizes the principles of privacy protection, data sovereignty, and fair use during data collection, storage, and utilization. It clarifies the boundaries and responsibilities of data processing, providing theoretical support for regulating the management of sensitive data in psychological crisis intervention.

2.2.2. Algorithmic Ethics Theory

Focuses on the fairness, transparency, and interpretability of algorithms, requiring algorithmic decision-making processes to avoid bias and discrimination, and safeguard the equal rights and interests of students from diverse groups. It offers ethical guidance for optimizing the design of AI intervention algorithms.

2.2.3. Educational Ethics Theory

Adheres to the inherent educational nature of nurturing individuals, emphasizing that technological applications must align with the fundamental goal of fostering virtue through education, safeguard students' subjectivity and human dignity, and avoid the impact of technological alienation on the essence of education.

2.2.4. Risk Governance Theory

Advocates the collaborative participation of multiple subjects in risk prevention and control, and constructs a full-chain governance system including pre-prevention, in-process supervision, and post-accountability. It provides methodological guidance for integrating the efforts of colleges and universities, technology providers, teachers, students, and other stakeholders in addressing ethical risks.

3. Identification of Ethical Risks of AI Technology in Psychological Crisis Intervention

This chapter systematically identifies various potential ethical risks in the application of AI technology in college students' PCI from three core dimensions: technology, institution, and cognition.

3.1. Technical Level: Inherent Limitations of Weak AI and Insufficient Ethical Considerations

Current AI technology remains in the weak AI stage, with significant deficiencies in algorithm interpretability and data processing security—for instance, opaque algorithmic logic makes it difficult to trace the basis of risk assessment decisions, making it difficult to fully adapt to the complex scenarios and personalized needs of psychological crisis intervention, thus resulting in inherent technical flaws. Meanwhile, during the technology R&D process, R&D teams focus more on improving technical performance and efficiency, with insufficient anticipation of scenario-specific ethical risks in PCI. Ethical guidelines are not embedded into the entire technical design process, leading to inherent ethical vulnerabilities in the technology itself.

3.2. Institutional Level: Imperfect Ethical Governance System and Lack of Supervision

Existing laws and regulations related to data security and privacy protection cannot fully cover the special scenarios of AI application in psychological crisis intervention, leaving legal gaps in issues such as multi-source data collection, algorithmic fairness,

and accountability definition. The field of college mental health education has not yet established unified standards and ethical norms for AI technology applications. Different colleges and universities lack consistent guidance in terms of technology selection, data processing, and intervention procedures, resulting in widespread irregularities in practice. Additionally, the absence of specialized ethical supervision agencies for full-process oversight, coupled with simplistic and inefficient supervision methods, hinders the timely detection and prevention of ethical risks.

3.3. Cognitive Level: Weak Ethical Awareness of Subjects and Inadequate Protection of Informed Consent Rights

Some colleges and universities overemphasize the efficiency improvements brought by technology, treating technological applications as “a key innovation in practice” while neglecting ethical risk prevention and control. They lack systematic ethical management mechanisms and risk assessment systems. Frontline staff, such as counselors and psychological teachers, lack sufficient knowledge reserves related to AI ethics, limiting their ability to identify, assess, and respond to ethical risks in technological applications, thus making it difficult to effectively mitigate risks. Students have insufficient understanding of the technical principles, data usage methods, and potential risks of AI-based interventions. Colleges and universities have not established effective information disclosure and feedback mechanisms, preventing students from effectively exercising their right to informed consent and supervision.

4. Analysis of the Causes of Ethical Risks in AI-Powered Psychological Crisis Intervention

The ethical risks identified above are not isolated; they stem from the interaction of technical, institutional, cognitive, and practical factors, which are analyzed in depth below.

4.1. Technical Level: Inherent Limitations of Weak AI

4.1.1. Immature Technological Development

Current AI technology is in the weak AI stage, with significant shortcomings in algorithm interpretability and data processing security. It cannot fully adapt to the complex scenarios and personalized needs of psychological crisis intervention, leading to inherent technical flaws.

4.1.2. Insufficient Ethical Considerations in Technical Design

During the R&D of AI technology, R&D teams prioritize improvements in technical performance and efficiency, with inadequate prediction of ethical risks in PCI scenarios. Ethical guidelines are not integrated into the entire technical design process.

4.1.3. Deficiencies in Scenario Adaptability

Technology R&D has not fully integrated the professional characteristics and practical needs of college psychological crisis intervention, lacking consideration for the personalization and complexity of intervention scenarios—such as ignoring cultural differences among students from ethnic minorities or cross-border

backgrounds—resulting in inherent obstacles to the practical application of technology.

4.2. Institutional Level: Imperfect Ethical Governance System

4.2.1. Lagging Laws and Regulations

Existing laws and regulations on data security and privacy protection fail to fully cover the special scenarios of AI in psychological crisis intervention, resulting in legal gaps in multi-source data collection, algorithmic fairness, and accountability definition.

4.2.2. Lack of Industry Norms

The field of college mental health education has not yet formulated unified standards and ethical norms for AI technology applications. Different colleges and universities lack consistent guidance in technology selection, data processing, and intervention procedures, leading to widespread irregularities in practice.

4.2.3. Inadequate Supervision Mechanisms

There is a lack of specialized ethical supervision agencies for full-process oversight of AI-powered psychological crisis intervention. Supervision methods are simplistic and inefficient, making it difficult to timely detect and effectively control ethical risks.

4.3. Cognitive Level: Weak Ethical Awareness of Subjects

4.3.1. Insufficient Ethical Attention from Colleges and Universities

Some colleges and universities overemphasize the efficiency gains from technology, treating technological applications as a “highlight” of work innovation while neglecting ethical risk prevention and control. They lack systematic ethical management mechanisms and risk assessment systems.

4.3.2. Inadequate Capabilities of Technology Users

Frontline staff, such as counselors and psychological teachers, lack sufficient knowledge of AI ethics, limiting their ability to identify, assess, and respond to ethical risks in technological applications, thus hindering effective risk mitigation in practice.

4.3.3. Inadequate Protection of Students' Right to Informed Consent

Students have insufficient understanding of the technical principles, data usage methods, and potential risks of AI-based interventions. Colleges and universities have not established effective information disclosure and feedback mechanisms—such as failing to provide accessible written explanations of data usage rules—preventing students from effectively exercising their right to informed consent and supervision.

4.4. Practical Level: Lack of Multi-Stakeholder Collaboration Mechanisms

4.4.1. Disconnection between Technology R&D and Educational Practice

There is a lack of in-depth communication and collaboration between AI technology providers and colleges and universities. The R&D of technical products does not fully integrate the professional needs and ethical requirements of

psychological crisis intervention, resulting in inadequate adaptability of products to practical scenarios.

4.4.2. Unclear Ethical Responsibilities in Home-School-Hospital Collaboration

In joint emergency interventions, there is a lack of clear ethical agreements and collaboration mechanisms among colleges and universities, families, and hospitals. The scope, methods, and privacy protection requirements for data sharing are ambiguous, which may easily lead to responsibility shirking or infringements.

4.4.3. Absence of Multi-Subject Linkage Mechanisms

Relevant subjects such as colleges and universities, technology R&D parties, and regulatory authorities lack normalized communication and linkage mechanisms, making it difficult to form synergy in risk prevention and control, problem disposal, and accountability, thus greatly reducing the effectiveness of risk prevention and control.

5. Construction of Ethical Risk Mitigation Mechanisms for AI Technology in Psychological Crisis Intervention

Based on the causes of ethical risks in four dimensions—technology, institution, cognition, and practice—this chapter constructs a comprehensive, multi-dimensional, and systematic ethical risk mitigation mechanism. Different from existing single-dimensional governance schemes, this mechanism integrates technical optimization, institutional constraints, ethical guidance, and practical collaboration, realizing the organic unity of technological empowerment and ethical integrity.

5.1. Technical Dimension: Optimize Technical Design and Fortify Ethical Defenses

Construct a privacy-protection-oriented technical system: Adopt technical measures such as data desensitization, encrypted storage, and hierarchical permission management to ensure the full-lifecycle security of psychological data; strictly adhere to the “minimum necessary” principle for data collection, clarify the scope and boundaries of data collection, and avoid over-collection of students’ sensitive information.

Enhance the fairness and interpretability of algorithms: Optimize the training data of algorithm models by incorporating diverse samples covering students of different genders, regions, and family backgrounds to reduce algorithmic bias; introduce technologies such as SHAP (SHapley Additive exPlanations) values and knowledge graphs to enhance the interpretability of algorithmic decision-making, and clearly present the basis for risk assessment and the logic of intervention plan matching to teachers and students.

Strengthen technical safety testing and iteration: Establish a safety assessment mechanism for AI intervention systems, conduct regular vulnerability detection and risk assessment, and promptly address security loopholes; continuously optimize technical algorithms and functional designs based on practical feedback and ethical requirements to improve the alignment between technology and ethical standards.

5.2. Institutional Dimension: Improve the Governance System and Clarify Ethical Norms

Formulate specialized ethical guidelines and technical standards: Based on the actual conditions of colleges and universities, issue the Ethical Norms for AI-Empowered Psychological Crisis Intervention, clarifying ethical requirements for data processing, algorithm application, and intervention implementation; develop technical application standards for AI-based psychological interventions, regulating the entire process of technology selection, system deployment, and effect evaluation.

Improve accountability division and liability mechanisms: Clarify the ethical responsibilities of colleges and universities, technology providers, psychological teachers, counselors, and other stakeholders, and sign responsibility agreements and checklists; establish an accountability system for ethical violations, and investigate the liabilities of relevant parties in accordance with the law for acts such as data breaches, algorithmic discrimination, and improper interventions, forming an effective constraint mechanism.

Construct a multi-stakeholder collaborative supervision mechanism: Establish an ethical supervision committee composed of ethical experts, psychological scholars, AI technology specialists, and teacher-student representatives to conduct full-process ethical supervision; leverage technologies such as blockchain to build an intervention process traceability system, enabling the full recording of data flow, algorithmic decision-making, and intervention behaviors to ensure traceable and verifiable supervision.

5.3. Ethical Dimension: Strengthen Awareness Cultivation and Build Consensus on Values

Enhance the ethical management awareness of colleges and universities: Incorporate AI ethics into the overall planning of college mental health work, establish an ethical risk prevention and control responsibility system, and clarify the management responsibilities at the school and college levels; strengthen the ethical review of AI intervention projects, and prohibit the implementation of projects without ethical evaluation.

Strengthen ethical training for relevant personnel: Conduct special training on AI ethics and data security for counselors and psychological teachers to improve their ability to identify and respond to risks; provide training on mental health professional knowledge and educational ethics for AI technology R&D personnel to promote the in-depth integration of technology R&D and ethical requirements.

Protect students' right to informed consent and participation: Fully inform students of the technical principles, data usage methods, and potential risks of AI-based interventions, and obtain their explicit informed consent in writing; establish student feedback mechanisms and ethical appeal channels, allowing students to raise objections and suggestions regarding ethical issues during the intervention process.

5.4. Practical Dimension: Optimize Application Scenarios and Strengthen Collaborative Linkages

Adhere to the principle of technical assistance rather than replacement: Clarify the auxiliary role of AI in psychological crisis intervention — key links such as core decision-making and emergency interventions must be led by professional psychological teachers; retain manual intervention channels to allow students to

independently choose intervention methods, avoiding technical alienation and the weakening of humanistic care.

Optimize personalized intervention scenarios: Develop a library of personalized intervention plans based on factors such as students' individual differences and cultural backgrounds to avoid a "one-size-fits-all" intervention model; rationally utilize new technologies such as VR meditation and metaverse group counseling to enhance the interest and acceptability of interventions while integrating humanistic care and emotional support.

Improve the ethical mechanism for home-school-hospital collaboration: Establish an ethical agreement for joint home-school-hospital interventions, clarifying the scope, methods, and privacy protection requirements for psychological data sharing; formulate ethical handling procedures for emergency crisis events to maximize the protection of students' privacy while ensuring their safety, and avoid ethical conflicts.

6. Conclusions and Prospects

This section summarizes the core findings of the research, condenses the key conclusions on ethical risk prevention and control in AI-empowered college students' psychological crisis intervention, and provides references for the sustainable development of this field.

6.1. Research Conclusions

AI technology offers significant advantages in improving the accuracy and efficiency of college psychological crisis intervention, but it also poses ethical risks related to data privacy and algorithmic fairness. These risks arise from the combined effects of technical limitations, institutional deficiencies, weak cognitive awareness, and insufficient collaboration. By constructing a "technology-institution-ethics-practice" four-in-one mitigation mechanism, ethical risks can be effectively prevented and resolved, achieving a dynamic balance between technological empowerment and ethical constraints, and enhancing the standardization and sustainability of intervention work.

6.2. Research Prospects

Future research can be expanded in three directions: First, conduct empirical research to verify the effectiveness and operability of the proposed mitigation mechanism in practical applications. Second, explore the differentiated governance of ethical risks in AI-enabled PCI under different cultural contexts to improve the cross-cultural adaptability of the mechanism. Third, keep pace with the development of generative AI and other emerging technologies, and iteratively optimize the mitigation mechanism to address new ethical challenges such as deepfake-related misinformation in psychological intervention.

Conflicts of Interest

The author declares that there are no potential or actual conflicts of interest related to the publication of this article.

Funding

This work was supported by the 2025 University-Level Research Project of Zhejiang Dongfang Polytechnic, entitled Research on the Path of AI-Enabled Identification and Intervention of College Students' Psychological Crisis (Grant No.: DF2025AIGC04).

References

- [1] Cong, C. Discussion on AI-Empowered Discourse Practice of Online Ideological and Political Education. *Ideological and Theoretical Education*, 2023, 3, 91-96.
- [2] Zhang, N.; Gao, P.C. Ethical Risks and Governance Paths of Generative Artificial Intelligence's Emotional Simulation: An Analysis Based on the Theoretical Framework of Techno-Social Co-construction. *Scientific Decision-Making*, 2025, 2, 123-134.
- [3] Peng, L.; Guo, L.; Li, Q.Y. Research on College Students' Mental Health Education from the Perspective of Generative Artificial Intelligence. *Journal of Foshan University*, 2024, 7, 97.
- [4] Liu, D.Y.; Luo, F.; Tu, Z.R.; Rao, S.J.; Shen, Y. Current Situation and Challenges of Artificial Intelligence Technology Empowering Psychological Development. *Journal of Beijing Normal University (Natural Science Edition)*, 2024, 60(1), 30-35.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)