Cognitive and Behavioral Science Digest
WISDOM ACADEMIC https://doi.org/10.63808/cbsd.v2i2.468 Volume 2 , Issue 2

Review Article

Artificial Intelligence-Enabled Mental Health Education in Primary and Secondary Schools: A Human-Centered Governance Framework

Yu Xie[1,*], Zhifang Guo[1]
[1]School of Teacher Education, Shangrao Normal University, Shangrao 334001, China
[*]**Correspondence:** Yu Xie, 2202310054@stu.jxufe.edu.cn

© The Author(s) 2026.
This article is published by Wisdom Academic Press Ltd. under a Creative Commons Attribution 4.0 International License. The license permits use, sharing, adaptation, distribution, and reproduction in any medium or format, provided that appropriate credit is given to the original author(s) and the source, a link to the license is supplied, and any modifications are indicated. Unless otherwise stated, all images and third-party materials included in this article are also covered by the same license. For any material not covered by this license, if the intended use is neither permitted by applicable law nor allowed under the license terms, direct permission must be obtained from the copyright holder. To view a copy of this license, please visit http://creativecommons.org/licenses/by/4.0/.

Abstract

This conceptual article examines how artificial intelligence (AI) may support mental health education in primary and secondary schools without assuming clinical authority. Using a targeted synthesis of recent peer-reviewed reviews, international guidance, Chinese policy and applicable legislation, it identifies four governance risks: psychological datafication, algorithmic misclassification, anthropomorphic dependence and digital inequity. It then develops five linked principles: developmental appropriateness, human review, tiered support boundaries, institutional accountability and multi-stakeholder collaboration. The framework's contribution lies in connecting each risk to permitted school functions, accountable actors, data limits, escalation procedures, monitoring indicators and withdrawal rules. AI may organize provisional signals, deliver low-threshold psychoeducation and support continuity, but it should not diagnose, assign final psychological risk or provide independent therapy. The framework is calibrated primarily for Chinese primary and secondary schools; transfer to other settings requires adjustment to local law, staffing and referral systems. Its propositions require participatory and empirical validation before claims of effectiveness can be made.

Keywords

Artificial intelligence, School mental health education, Adolescent mental health, Conversational AI, Data governance, Human-centered intervention

1. Introduction

Adolescence is a developmental period in which cognitive control, emotion regulation, identity exploration and peer relations change rapidly. The World Health Organization (2025) reports that one in seven 10- to 19-year-olds experiences a mental disorder. In China, recent policy has placed student

mental health in a more prominent position and has encouraged AI-based psychological assistants and intelligent stress-relief spaces under appropriate safeguards (Ministry of Education of the People's Republic of China, 2025).

AI may help schools organize weak signals, provide basic psychoeducation and maintain follow-up communication when professional staff are limited (Yu & Zhang, 2023, 2024). The evidence does not support a general claim of effectiveness, however. Independent reviews of digital mental health interventions for young people report modest or heterogeneous effects, stronger engagement when a human component is retained, and little evidence on cost-effectiveness or transfer to low-resource settings (Lehtimaki et al., 2021; Liverpool et al., 2025). These limits matter because privacy intrusion, biased prediction, overreliance on conversational systems and unequal service capacity are especially consequential for minors (Guo & Wu, 2023; Wang, 2025).

The question addressed here is therefore not whether schools should use AI, but under what institutional boundaries AI may support mental health education without assuming clinical authority. The framework is calibrated primarily for Chinese primary and secondary schools, where national education policy and data-protection law determine institutional duties. International guidance is used to identify broader child-rights and health-governance principles; applying the framework elsewhere would require recalibration to local law, school staffing and referral systems. The paper contributes a school-specific linkage among four risk mechanisms, five governance principles, human decision rights and tiered support boundaries rather than proposing a new set of general AI ethics principles.

2. Conceptual Method and Theoretical Positioning

2.1. Source selection and framework development

This article uses a targeted conceptual synthesis, not a systematic review. A structured search was conducted in August 2026 in PubMed and ERIC for English-language material published from January 2020 through August 2026. Search combinations joined terms for technology ("artificial intelligence," "digital mental health," "conversational AI"), population ("child," "adolescent," "student"), setting ("school," "education") and issue ("mental health," "psychoeducation," "governance," "privacy" or "ethics"). Recent reviews were followed backward to relevant earlier work. Official repositories of WHO, UNESCO, UNICEF and Chinese government bodies were searched for guidance, professional standards, education policy and legislation; Crossref and publisher records were used to verify bibliographic metadata.

Sources were included when they addressed school-aged minors, digital or AI-enabled mental health functions, child-centered governance, school mental health practice or data protection. Adult-only clinical studies, product descriptions and sources without a clear bearing on educational governance were excluded, although health-governance guidance was retained where it clarified the

boundary between educational support and clinical care. Search results were reviewed iteratively rather than through an exhaustive PRISMA process, and no effect estimates were calculated.

Framework development proceeded through problem-to-control coding. Recurring concerns were grouped according to what could go wrong in a school: pervasive psychological data collection, erroneous inference, relational displacement and unequal implementation capacity. Governance responses were then clustered by the decision they constrain: developmental fit, human review, support-tier boundaries, institutional accountability and collaboration. The provisional categories were checked against empirical reviews, child-rights guidance, Chinese professional standards and legislation. This derivation does not establish empirical validity.

2.2. Cognitive and behavioral governance

Cognitive and behavioral governance refers here to institutional rules that shape how AI-generated information changes attention, interpretation and action in a school. An alert can redirect a teacher's attention; a risk label can alter later expectations; a conversational response can affect a student's understanding of distress and willingness to seek human help. The object of governance is therefore not only the model but also the pathway through which an output becomes a judgment, record or intervention.

This lens is narrower than responsible AI governance and more operational than a general statement of human-centered design. UNICEF guidance foregrounds children's rights and best interests, UNESCO addresses the educational use of generative AI, and WHO guidance assigns accountability and human control in health settings (UNESCO, 2023; UNICEF Innocenti, 2025; World Health Organization, 2021). The present framework brings those concerns into school mental health education by connecting specific risks to role assignments, support tiers, data limits and exit rules. Its principles are not novel in isolation. The distinctive claim lies in their combination around the cognitive and behavioral consequences of school decisions.

Developmental and epistemic limits anchor the analysis. Children and adolescents are still developing emotional vocabulary, social judgment and help-seeking capacity, so the autonomy given to a system and the point of human transfer must vary with age and context. AI outputs are probabilistic and sensitive to training data, prompts and behavioral traces. Ordinary examination pressure, family circumstances and serious distress may produce similar observable patterns. The outputs are provisional signals, not measurements of an underlying psychological state; interpretation and follow-up remain human responsibilities (Chinese Psychological Society, School Mental Health Education Professional Committee, 2026).

3. Functional Boundaries of AI in School Mental Health Education

3.1. Preventive signal organization and review

Student distress may appear as changes in attendance, learning engagement, peer interaction, emotional tone or help-seeking behavior. AI can organize such information and identify patterns for review, but it cannot determine that a student is psychologically at risk. Validation remains uneven across populations and contexts, and generative systems may produce unstable or biased responses (Mu et al., 2026; Sharma et al., 2025). An alert should open a human inquiry rather than close it with a classification.

3.2. Low-threshold psychoeducation and self-regulation support

Digital tools can offer mental health literacy, emotion naming, mood journaling, breathing exercises and help-seeking prompts to students who hesitate to seek face-to-face support. Reviews report promising but heterogeneous outcomes, with stronger support for structured interventions and human-guided components than for fully automated systems (Fernández-Batanero et al., 2025; Hong et al., 2025; Lehtimaki et al., 2021). In a school, these functions should remain educational. A conversational system may explain a coping exercise or show how to contact a counselor; it should not present itself as a therapist or deliver independent clinical advice (Hua et al., 2025).

3.3. Continuity of educational support

Classroom education, counseling appointments, guardian communication and external referral are often disconnected. An AI-supported platform may organize reminders, non-sensitive follow-up records and resource lists. Personalization should be limited to a defined educational purpose, and continuity should not become comprehensive profiling. The relevant outcome is whether a student reaches accountable human support, not how long the student remains engaged with a platform.

4. Risk Mechanisms

4.1. Psychological datafication

Mental health data are highly sensitive because they may involve emotions, family conflict, counseling notes, self-harm risk and crisis records. When such data are collected at scale, students may be transformed from developing persons into objects of continuous measurement. The danger is not only data leakage; it is also over-collection. If schools collect classroom behavior, online learning traces and social interaction data without strict limits, mental health education may become psychological surveillance.

Data collection should therefore begin with a named support purpose and a minimum-data test. Psychological information must not be repurposed for ranking, discipline, advertising or unrelated academic evaluation. Access, retention and deletion rules are part of the educational boundary because records can continue to shape adult expectations after the original context has passed.

4.2. Algorithmic misclassification and label fixation

The same behavior may reflect ordinary examination stress, family conflict, peer exclusion, neurodevelopmental difference or serious distress. Narrow training samples and static indicators can produce false positives that stigmatize students and false negatives that delay attention. Once a student is marked “high risk,” later behavior may be read through the label. Results must be reviewable and reversible, and an unconfirmed alert should not enter a permanent student record.

4.3. Anthropomorphic dependence

Conversational AI may appear patient, non-judgmental and always available. These features can encourage self-disclosure but may also lead emotionally vulnerable adolescents to rely on machines for attachment, validation and decision-making. Reviews of conversational AI in pediatric mental health emphasize its promise but also the need for careful safety evaluation, escalation protocols and human supervision (Mansoor et al., 2025).

The deeper risk is relational displacement. If AI systems become the primary source of emotional support, schools may unintentionally weaken opportunities for students to build real trust with teachers, parents, counselors and peers. Human relationships are not an inefficient substitute for AI; they are the core medium through which school mental health education works.

4.4. Digital inequity

Implementation capacity changes the meaning of the same tool. A school with trained counselors, secure infrastructure and an external referral network may use an alert as one input to professional review. In a school without those resources, the alert may remain unreviewed or be treated as a conclusion. Evidence on cost-effectiveness and low-resource generalizability is still sparse (Lehtimaki et al., 2021; Liverpool et al., 2025). Equity cannot be inferred from access to software alone; it also depends on staffing, review time, language and cultural fit, secure data handling and an offline route to assistance.

5. A Human-Centered Governance Framework

The framework assigns five principles to decisions made before, during and after deployment. Developmental appropriateness controls design and communication; human review controls interpretation; tiered boundaries limit permitted functions; accountability governs data, vendors and monitoring; collaboration identifies who must act. Table 1 maps these principles to the four risks and gives a compact operational pathway.

Developmental appropriateness begins with age-sensitive explanations of purpose, uncertainty and data use. In China, the Personal Information Protection Law treats medical-health information and

the personal information of children under 14 as sensitive; it requires a specific purpose, strict necessity, separate consent for sensitive information and guardian consent plus special processing rules for children under 14 (Standing Committee of the National People's Congress, 2021b). The Law on the Protection of Minors likewise requires lawful, legitimate and necessary processing and provides for guardian consent for children under 14 and correction or deletion rights (Standing Committee of the National People's Congress, 2021a). These legal requirements do not exhaust good practice. Students should receive an age-appropriate explanation and a meaningful opportunity to assent or decline optional educational functions. Guardian access should be limited when unrestricted disclosure of a student's private reflection would undermine safety, trust or the student's developing autonomy; safeguarding duties and applicable law may nevertheless require disclosure in a credible emergency.

Human review is a distribution of responsibility, not the presence of any adult near the system. Teachers may notice a change, support classroom activities and initiate referral, but they should not interpret psychological risk scores. School counselors review alerts, speak with the student, document the basis for action and decide whether the matter remains educational or requires referral. School leaders approve vendors, staffing and incident procedures. Guardians provide contextual information and participate when consent or safeguarding rules require it. Clinical assessment belongs to qualified external mental-health professionals. Vendors remain responsible for security, documentation and correction of system defects; they do not decide what happens to an individual student.

Tier movement must not be controlled by a fixed score. Universal use is limited to general literacy, self-regulation exercises and help-seeking information without individualized psychological inference. Targeted support begins after a student request or a human-observed concern and only when a counselor defines the purpose, information needed and review date. Content suggesting immediate danger triggers an urgent human safeguarding pathway: the system stops routine dialogue, displays human contact options and notifies the designated professional under an approved protocol. The conditions are set by qualified school mental-health staff with reference to local emergency procedures, then tested and reviewed. False positives are documented and corrected; possible false negatives are examined through incident review, comparison with human referrals and accessible non-digital reporting channels.

Accountability begins before procurement. A school should document the problem, the least data-intensive option, decision rights, retention period and vendor obligations. The Regulations on the Protection of Minors in Cyberspace add duties concerning minors' personal information and protective responses to private information online (State Council of the People's Republic of China, 2023). Contracts should prohibit secondary commercial use and unrelated educational reuse. Students and guardians need routes to access, correct or delete information and challenge an outcome. Monitoring should cover false alerts, missed incidents, response time, group disparities, follow-up and complaints.

WISDOM ACADEMIC https://doi.org/10.63808/cbsd.v2i2.468 Volume 2 , Issue 2

Material misrepresentation, repeated escalation failure, unlawful reuse, uncorrected bias or review demand beyond available staff capacity is a reason to suspend or withdraw the tool.

Table 1

Operational mapping of risks, principles, actors and controls

Risk mechanism	Principles and accountable actors	Permitted function and data limit	Human review and escalation	Monitoring and stop rule
Psychological datafication	Developmental appropriateness; accountability. School leader, data steward, counselor, vendor.	Named educational purpose; minimum data; no ranking, discipline or secondary use.	Age-appropriate notice and assent; consent; limited access; correction/deletion route.	Audit access, retention and complaints; suspend for unlawful reuse or breach.
Algorithmic misclassification	Human review; tiered boundaries. Counselor, teacher, external professional.	AI organizes signals but does not diagnose or assign final risk; unconfirmed alerts are not permanent records.	Counselor checks context; urgent matters enter the approved human safeguarding pathway.	Track errors, response time and group disparities; suspend for poor validity, bias or excessive review burden.
Anthropomorphic dependence	Developmental appropriateness; collaboration; tiered boundaries. Counselor, teacher, guardian, vendor.	Psychoeducation, coping prompts and navigation only; no therapist representation or independent treatment.	Visible AI identity; urgent content stops routine dialogue and transfers to a professional.	Review dependence, unsafe advice and escalation; withdraw for misrepresentation or repeated failure.
Digital inequity	Accountability; collaboration. School leader, local authority, vendor, community provider.	Low-data universal functions; no individualized alerts without timely human review.	Named human contact, offline access and referral route; shared regional support where feasible.	Track access, language fit, cost and referral completion; narrow or withdraw if gaps persist.

Collaboration and feasibility determine which functions a school may responsibly offer. Under-resourced schools should begin with a minimum-governance model: general psychoeducation and resource navigation that collect no psychological profile, combined with a named human contact and an external referral directory. Individualized alerting should not be introduced where timely counselor review cannot be guaranteed. County-level or regional counselor pools, shared procurement and referral agreements may distribute expertise, but they do not remove the need to budget for review time, staff training, security assessment and vendor exit. A narrower tool with a functioning human pathway is preferable to a broad system whose alerts cannot be acted upon.

6. Discussion

Responsible AI, child-centered guidance and health-AI ethics already establish safety, privacy,

human control and accountability. What requires separate treatment in schools is the chain from an uncertain output to attention, labeling, family communication and referral. The four-by-five mapping identifies where that chain can fail and who has authority to interrupt it. Tiered boundaries add a functional test: the more a tool moves from general education toward individualized psychological inference, the stronger the professional, legal and evidential requirements become.

This distinction also calibrates the evidence base. Digital interventions may improve some outcomes, especially when they use established content and retain human guidance, but results vary and long-term, cost and low-resource evidence remain limited (Lehtimaki et al., 2021; Liverpool et al., 2025). Those findings support cautious testing of bounded functions; they do not establish that automated screening or conversational systems are effective school mental health services. The governance framework is therefore a conceptual proposal for controlling implementation, not evidence that implementation will produce benefit.

6.1. Practical implications

School leaders should start with a defined educational need and a time-limited pilot. Procurement review should involve leadership, counseling staff, information-security personnel and, where feasible, student and guardian representatives. Before launch, the school should test ambiguous and high-risk scenarios, confirm who receives an alert and set a maximum review time. Training should make uncertainty visible: teachers need to know what an alert cannot establish, counselors need authority to override it, and students need to know that the system may be wrong and does not replace a trusted person.

Evaluation should focus on the support pathway rather than user volume. Relevant questions include whether students reach a person sooner, whether human follow-up occurs, whether some groups receive disproportionate alerts and whether correction requests are resolved. Staff time spent reviewing low-value alerts is a real cost. If review capacity, secure storage or referral access is absent, the permitted function should be narrowed rather than responsibility shifted to an automated system.

6.2. Limitations and future research

The targeted search was designed to support conceptual development and may have missed relevant studies, particularly non-English evidence and work outside the selected databases. The underlying literature covers different technologies, age groups and clinical or educational settings; broad transfer to Chinese schools is uncertain. No students, guardians, teachers or counselors were involved in deriving the framework, and no empirical test shows that the four risks are exhaustive, that the five principles are sufficient or that the proposed matrix improves decisions.

Validation should examine implementation as well as comprehension. Pilot studies could test whether responsibilities are followed, escalation procedures work under time pressure, inaccurate

records are corrected and monitoring detects unequal effects. Comparisons across rural, urban and under-resourced schools should include staffing and lifecycle cost, not only technical performance. Participatory work is also needed to examine conflicts between student privacy, guardian involvement and safeguarding, including situations in which automated advice conflicts with advice from a trusted adult.

7. Conclusion

AI may support psychoeducation, limited signal organization and continuity of educational support, but it cannot independently diagnose, assign psychological risk or provide therapy. A defensible school model links developmental fit, human decision rights, tiered functions, data governance and collaboration to explicit monitoring and withdrawal rules. In Chinese schools these controls must operate within national data and minors-protection law; elsewhere they require local legal and institutional calibration. Responsibility for a minor's mental health remains with accountable people and institutions.

Acknowledgments

Not applicable.

Author contributions

Conceptualization, Y.X.; methodology, Y.X.; writing—original draft preparation, Y.X.; writing—review and editing, Y.X. and Z.G.; supervision, Z.G. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Ethics approval

Ethical approval was not required for this study.

Consent to participate

Not applicable.

Consent for publication

Not applicable. This manuscript contains no identifiable personal data in any form.

Additional information

Received: 12 July 2026

Accepted: 1 September 2026

Published online: 17 November 2026

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Chinese Psychological Society, School Mental Health Education Professional Committee. (2026). Professional competence standards for school mental health teachers. *Acta Psychologica Sinica, 58*(1), 185-190. https://doi.org/10.3724/SP.J.1041.2026.0185

[2] Fernández-Batanero, J. M., Fernández-Cerero, J., Montenegro-Rueda, M., & Fernández-Cerero, D. (2025). Effectiveness of digital mental health interventions for children and adolescents. *Children, 12*(3), 353. https://doi.org/10.3390/children12030353

[3] Guo, S. N., & Wu, Y. H. (2023). How can child-centered artificial intelligence be realized? *Modern Educational Technology, 33*(11), 29-37. https://doi.org/10.3969/j.issn.1009-8097.2023.11.003

[4] Hong, S. H., Chun, T. K., Nam, Y. J., Kim, T. W., Cho, Y. H., Son, S. J., Roh, H. W., & Hong, C. H. (2025). Digital mental health interventions for adolescents: An integrative review based on the behavior change approach. *Children, 12*(6), 770. https://doi.org/10.3390/children12060770

[5] Hua, Y., Na, H., Li, Z., Liu, F., Fang, X., Clifton, D., & Torous, J. (2025). A scoping review of large language models for generative tasks in mental health care. *npj Digital Medicine*, 8, 230. https://doi.org/10.1038/s41746-025-01611-4

[6] Lehtimaki, S., Martic, J., Wahl, B., Foster, K. T., & Schwalbe, N. (2021). Evidence on digital mental health interventions for adolescents and young people: Systematic overview. *JMIR Mental Health, 8*(4), e25847. https://doi.org/10.2196/25847

[7] Liverpool, S., Mc Donagh, J. E., Feather, J., Uzondu, C., Howarth, E., Bannerman, S., Kaehne, A., Foster, D., & Mateus, C. (2025). Updates on digital mental health interventions for children and young people: Systematic overview of reviews. *European Child & Adolescent Psychiatry, 34*(10), 2961-2974. https://doi.org/10.1007/s00787-025-02722-9

[8] Mansoor, M., Hamide, A., & Tran, T. (2025). Conversational AI in pediatric mental health: A narrative review. *Children, 12*(3), 359. https://doi.org/10.3390/children12030359

[9] Ministry of Education of the People's Republic of China. (2025). Ten measures for strengthening mental health work for primary and secondary school students. https://www.moe.gov.cn/srcsite/A06/s3325/202510/t20251020_1417420.html

[10] Mu, Y., Li, Q., Wang, Z., Zhang, L., & Chen, Y. (2026). Semantic stability, response strategies and bias of generative AI in mental health education. *Journal of Psychological Science, 49*(2), 258-270. https://doi.org/10.16719/j.cnki.1671-6981.20260201

[11] Sharma, G., Yaffe, M. J., Ghadiri, P., Gandhi, R., Pinkham, L., Gore, G., & Abbasgholizadeh-Rahimi, S. (2025). Use of artificial intelligence in adolescents' mental health care: Systematic scoping review of current applications and future directions. *JMIR Mental Health*, 12, e70438. https://doi.org/10.2196/70438

[12] Standing Committee of the National People's Congress. (2021a). Law of the People's Republic of China on the Protection of Minors. https://www.moe.gov.cn/jyb_sjzl/sjzl_zcfg/zcfg_qtxgfl/202110/t20211025_574798.html

[13] Standing Committee of the National People's Congress. (2021b). Personal Information Protection Law of the People's Republic of China. https://www.gjbmj.gov.cn/n1/2021/0903/c409088-32217039.html

[14] State Council of the People's Republic of China. (2023). Regulations on the Protection of Minors in Cyberspace (State Council Decree No. 766). https://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202310/t20231025_1087333.html

[15] UNESCO. (2023). Guidance for generative AI in education and research. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

[16] UNICEF Innocenti. (2025). Guidance on AI and children: Version 3.0. https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

[17] Wang, Y. Y. (2025). AI-enabled adolescent mental health education: Opportunities, challenges and responses. *People's Tribune, 2025*(7), 42-45. https://paper.people.com.cn/rmlt/pc/content/202504/17/content_30080933.html

[18] World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. https://www.who.int/publications/i/item/9789240029200

[19] World Health Organization. (2025). Mental health of adolescents. https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

[20] Yu, G. L., & Zhang, Z. (2023). Digital technology-enabled school mental health services. *Tsinghua Journal of Education, 44*(1), 19-29. https://tje.ioe.tsinghua.edu.cn/oa/darticle.aspx?id=202301004&type=view

[21] Yu, G. L., & Zhang, Z. (2024). Digital technology-enabled responses to student psychological crises. *Tsinghua Journal of Education, 45*(4), 25-36. https://tje.ioe.tsinghua.edu.cn/oa/darticle.aspx?id=202404003&type=view