

23 - 25 October 2026

Geneva , Switzerland

A Proposed Ethical Framework for AI-Based Digital Mental-Health Tools for Young People and Students

Hanan Namrouti , Cecilia Sik-Lányi

University of Pannonia, Hungary

Abstract

AI-based digital mental-health tools have the potential to improve access to mental-health information and support for young people, but their use also presents significant ethical and safety risks. This paper examines six key safeguards needed to ensure the responsible development and deployment of these technologies: transparency, human oversight, crisis escalation, privacy by design, relational safety, and continuous evaluation. Transparency requires young people to understand when they are interacting with AI and be aware of its capabilities and limitations. Human oversight ensures that AI complements rather than replaces qualified mental-health professionals, particularly in high-risk situations. Crisis escalation mechanisms are necessary to identify and respond appropriately to self-harm, suicide risk, abuse, and other serious concerns. Privacy by design protects sensitive mental-health information through data minimization, security, and transparent data practices. Relational safety addresses the risk of emotional dependency by preventing systems from encouraging secrecy, exclusivity, or replacement of human relationships. Finally, continuous evaluation ensures that risks such as bias, misinformation, inappropriate crisis responses, and harmful changes following system updates are identified and addressed over time. These safeguards reflect emerging international recommendations emphasizing evidence-based development, youth participation, accountability, crisis-referral mechanisms, and assessment of long-term effects. Together, they provide a framework for developing AI mental-health tools that prioritize young people's safety, autonomy, privacy, and wellbeing.

Keywords: Artificial Intelligence; Digital Mental Health; Young People; Ethical Safeguards; Mental Health.