Ashleigh Golden, PsyD

wayhaven

Evaluating AI-powered mental wellness tools

A practical guide for higher education decision-makers.

Empowering campuses to meet student needs with clarity, care, and confidence, even as resources tighten and expectations rise.

Introduction

Student needs continue to rise amid a persistent campus mental health crisis, while traditional support systems face ongoing staffing shortages, budget constraints, and political scrutiny. In this environment, colleges and universities are exploring new ways to promote student well-being, including generative artificial intelligence (AI) tools designed to support mental wellness.

Generative AI (GenAI) refers to artificial intelligence systems like ChatGPT that generate human-like text in response to user input. In mental wellness contexts, these tools are being explored for their potential to simulate conversation, offer emotional support, and encourage self-reflection. As student needs rise and digital behaviors shift, many campuses are asking whether and how these tools might support broader mental health and wellness strategies. Yet not all tools are created equal, and those marketed as "fit-for-purpose" can still vary widely in quality, safety, and alignment with student needs.

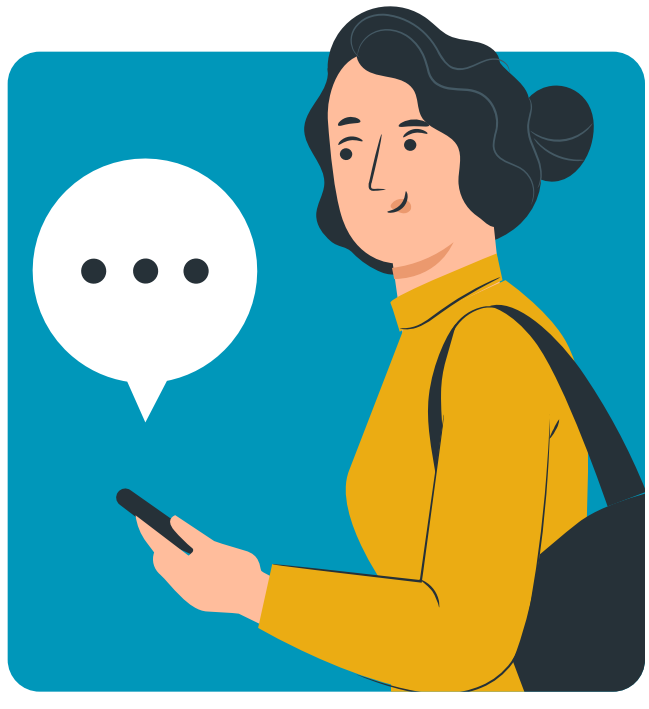
Interest in GenAI tools is rising, with sessions at events like the American College Health Association 2025 Annual Meeting — including *"Applications for AI in Student Health Today"* and *"Broadening Our Scope: Leveraging Generative AI for Enhanced Collegiate WellBeing"* — spotlighting how students are already turning to general-purpose and entertainment platforms (e.g., ChatGPT, Replika) for emotional support, revealing a growing gap between student use and the campus tools and structures available to support it safely and meaningfully. This signals a major behavioral shift: Students are already leveraging AI for mental wellness, whether or not institutions provide structured guidance.

Yet one core gap remains: *how* to evaluate these tools responsibly. In conversations with student affairs and counseling center leaders, one theme is clear: while interest is growing, most institutions lack a structured, values-aligned framework for assessing these technologies, even as expectations rise and they are asked to do more with less.

Campus leaders must navigate a fast-moving, fragmented marketplace while balancing the priorities of diverse stakeholders: students seeking accessible support, parents and faculty concerned about well-being and academic success, institutional leaders managing retention and risk, and counseling center staff who may be cautious but curious. Even as new GenAI tools emerge, many lack transparency around safety practices, blur the line between wellness and clinical care, or fail to demonstrate meaningful outcomes.

Without a structured approach to evaluation, institutions risk adopting tools that seem innovative but fall short on inclusivity, effectiveness, or student safety. This guide is designed to close that gap. To our knowledge, it is **the first practical framework created specifically to help higher education leaders evaluate GenAI tools for student mental wellness**. It outlines research-informed criteria to support safe, ethical, student-centered decision-making so that campuses can act with clarity, care, and confidence.

"Students are already leveraging AI for mental wellness, whether or not institutions provide structured guidance."

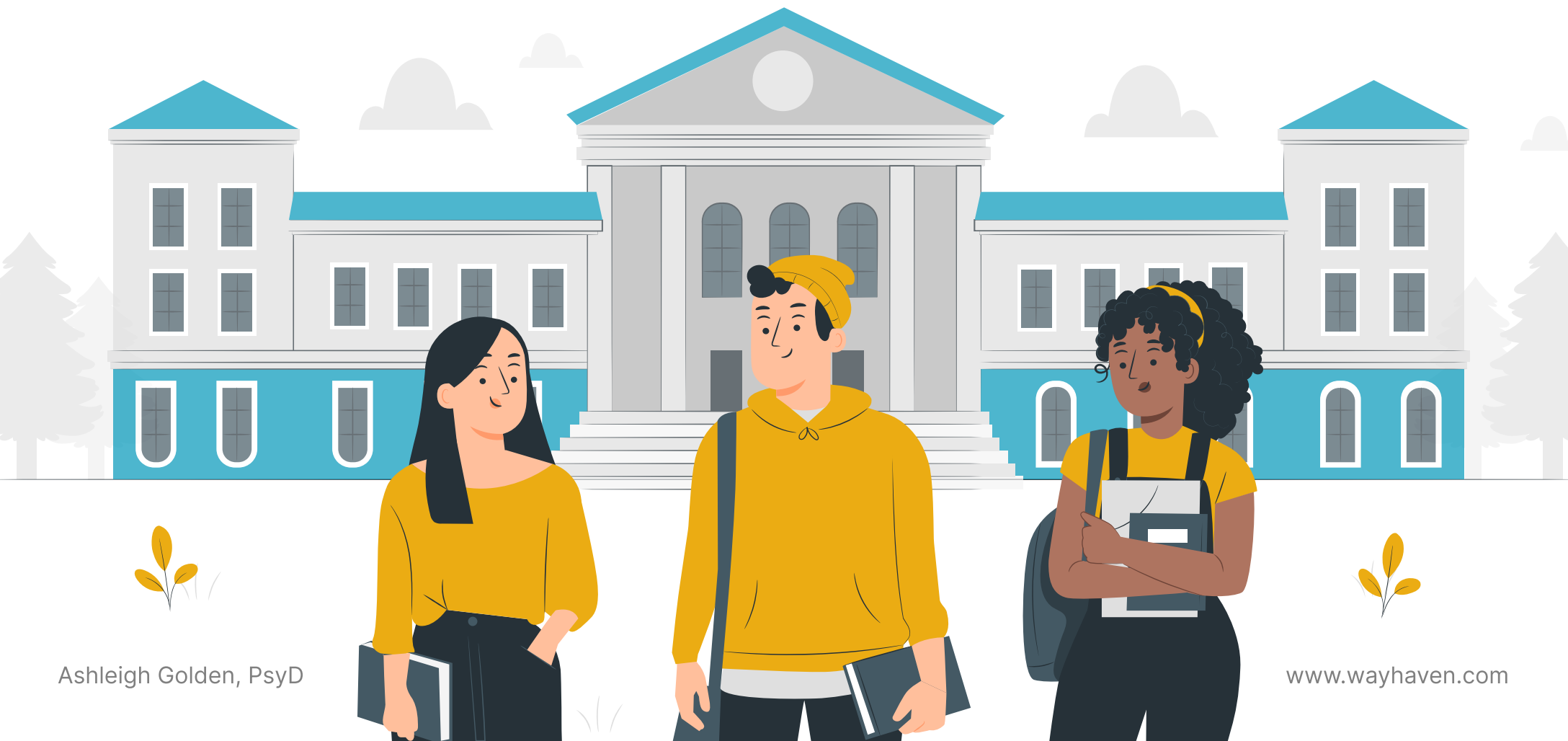


SECTION 1

How AI tools differ

Not all AI-powered tools serve the same purpose. Differences in design, safeguards, and scope can influence not only student mental wellness, but also help-seeking behavior, campus engagement, and institutional outcomes.

Tool type	Limitations	"Fit for purpose" alternatives
General purpose LLMs (e.g. ChatGPT, Anthropic, Gemini)	Not aligned with best safety practices, often reinforces users' emotions or thoughts without helping them reflect or reframe — even when those thoughts are distressing or unhelpful, inadequate psychological scaffolding, lacks crisis referral or meaningful guardrails	Fit-for-purpose mental wellness AI, trained on empirically supported frameworks and designed in ongoing collaboration with psychologists; supports emotion validation while gently encouraging reflection and adaptive action; aligned with college population needs and safety protocols and designed to support outcomes tied to well-being, engagement, and retention
Entertainment chatbots (e.g., Replika, Character.ai)	Risk of dependency, AI mimics emotional intimacy or attachment, no support for healthy real-life integration	Promote real-world social connection and belonging, encourages healthy behavior change and skills practice



SECTION 2

Not all “fit-for-purpose” tools meet the mark

As AI mental wellness solutions gain traction, some tools are now marketed as “fit-for-purpose” — yet still fall short of standards that define safe, effective, and developmentally appropriate support.

Common limitations include:

- Use of off-the-shelf LLMs with minimal evidence-informed guardrails
- Lack of grounding in empirically-supported methods
- Absence of clinical effectiveness data or outcome measures
- Weak or missing psychological scaffolding to support emotional and cognitive growth
- Prescriptive advice-giving that bypasses self-reflection and reinforces passivity
- Surface-level personalization that fails to align with individual user goals
- Overvalidation or sycophancy: uncritical agreement that may reinforce unhelpful beliefs and undermine growth
- Lack of inclusive design practices or representative user testing
- No integration of robust crisis protocols or connection to real-world resources that promote social engagement and safety
- Absence of clear, ongoing processes for iterative improvement based on real-world use

By contrast, **truly fit-for-purpose tools** are co-developed with psychologists and grounded in research-backed methods. They demonstrate measurable effectiveness, often through peer-reviewed publications or real-world usage data. These tools guide users to reflect and develop their own insights — rather than dispensing one-size-fits-all advice — and gently challenge unhelpful thoughts in ways that support positive cognitive and emotional shifts. Instead of fostering dependence, they promote real-world skill-building and behavior change.

High-integrity tools are also inclusive by design, reflecting diverse user identities and lived experiences. They use conversational memory and contextual awareness to personalize support, integrate robust real-time crisis protocols, and connect users to appropriate local resources that promote real-world engagement. Finally, they demonstrate a clear commitment to improvement, including continuous user testing, feedback loops, and alignment with the latest psychological science.

SECTION 3

Why “fit-for-purpose” AI tools matter

Not all mental wellness tools — whether powered by GenAI, delivered by humans, or something in between — are built with student needs in mind. Understanding the context in which a tool was designed is key to evaluating its value, limitations, and role on campus. The table below outlines how context of design directly impacts a tool’s strategic role on campus.

CONTEXT	STRATEGIC IMPLICATIONS
Students are already using GenAI like ChatGPT, Claude, or Replika for mental health support, but these tools were not built for safety, real-world behavior change, or student-specific needs.	Fit-for-purpose tools are trained on evidence-based content, grounded in ethical principles, and explicitly designed to promote mental wellness, teach coping skills, and promote healthy social connection — without risk of dependence.
Traditional and virtual care platforms, including teletherapy and peer support apps, remain dependent on clinician or peer availability, constrained by licensing, and may still feel stigmatizing or inaccessible to students from minoritized backgrounds or those facing other barriers to entry, such as social anxiety.	Purpose-built AI mental wellness tools can complement human systems by offering a more scalable, accessible entry point, especially for students who are unlikely to engage in traditional services as a first step or at all.
Students from historically marginalized or excluded groups may feel more comfortable opening up to an AI than to a human.	AI can reduce stigma and medical mistrust, offering a safer first step into care, particularly for students who avoid traditional services due to prior experiences of harm or lack of trust.
Many existing campus tools feel outdated or irrelevant to Gen Z.	Gen Z–friendly tools can communicate in familiar ways, operate 24/7, and provide personalized, campus-specific support that meets students where they are.
Staff hiring cannot keep pace with student demand — and mental health staffing may even be shrinking.	GenAI offers scalable, always-on support that can extend existing services, fill staffing gaps, and help during surge periods or with high-need populations.
Many institutions are integrating AI into administrative workflows, but not yet into student-facing mental wellness or care.	GenAI mental wellness tools fill this gap by offering safeguards, equity-informed design, and value that aligns with students’ lived experiences and institutional priorities.

SECTION 4

Evidence integrity, safety, and cultural inclusivity

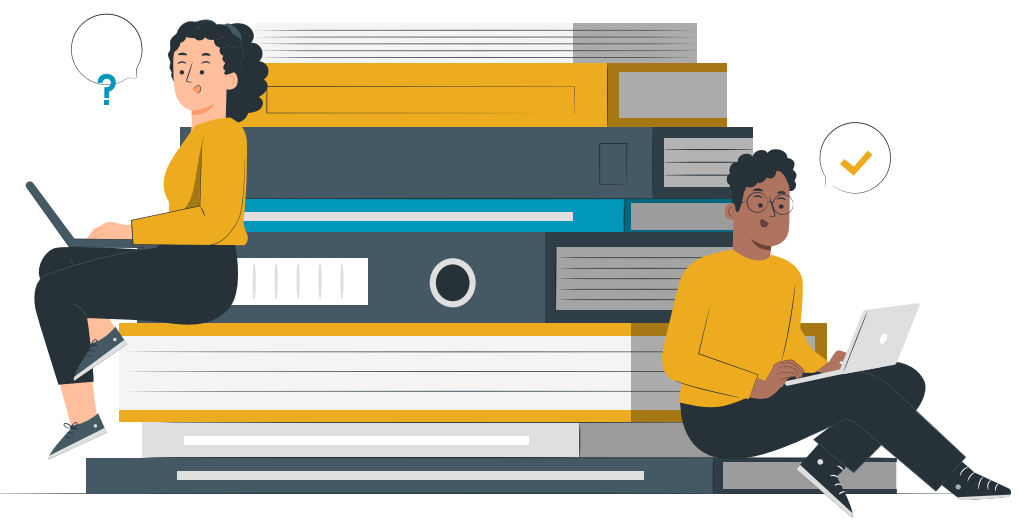
These questions can help institutions identify tools that are not only effective, but also safe, inclusive, and designed to foster autonomy, self-efficacy, and core life skills during the college years.

Privacy & consent

- Is user information kept private and confidential?
- Are privacy policies presented in accessible language?
- Is consent obtained before data collection?
- Do users have the ability to access data?
- Is it clear which outside entities, if any, data will be disclosed to, and under what conditions?

Safety & crisis management

- Are potentially dangerous or risky user behaviors (e.g., suicidal ideation, self-harm, violence) monitored?
- Does the tool include clear boundaries around what it can and cannot do?
- Does the tool redirect users to appropriate human care, including campus resources, when needed?
- Are there real-time, tailored crisis protocols with links to hyper-local emergency services?
- Are resources presented in a way that is optimized for utilization?
- Does the AI model emotionally healthy behavior rather than maladaptive traits (e.g., narcissism)?



Evidence integrity

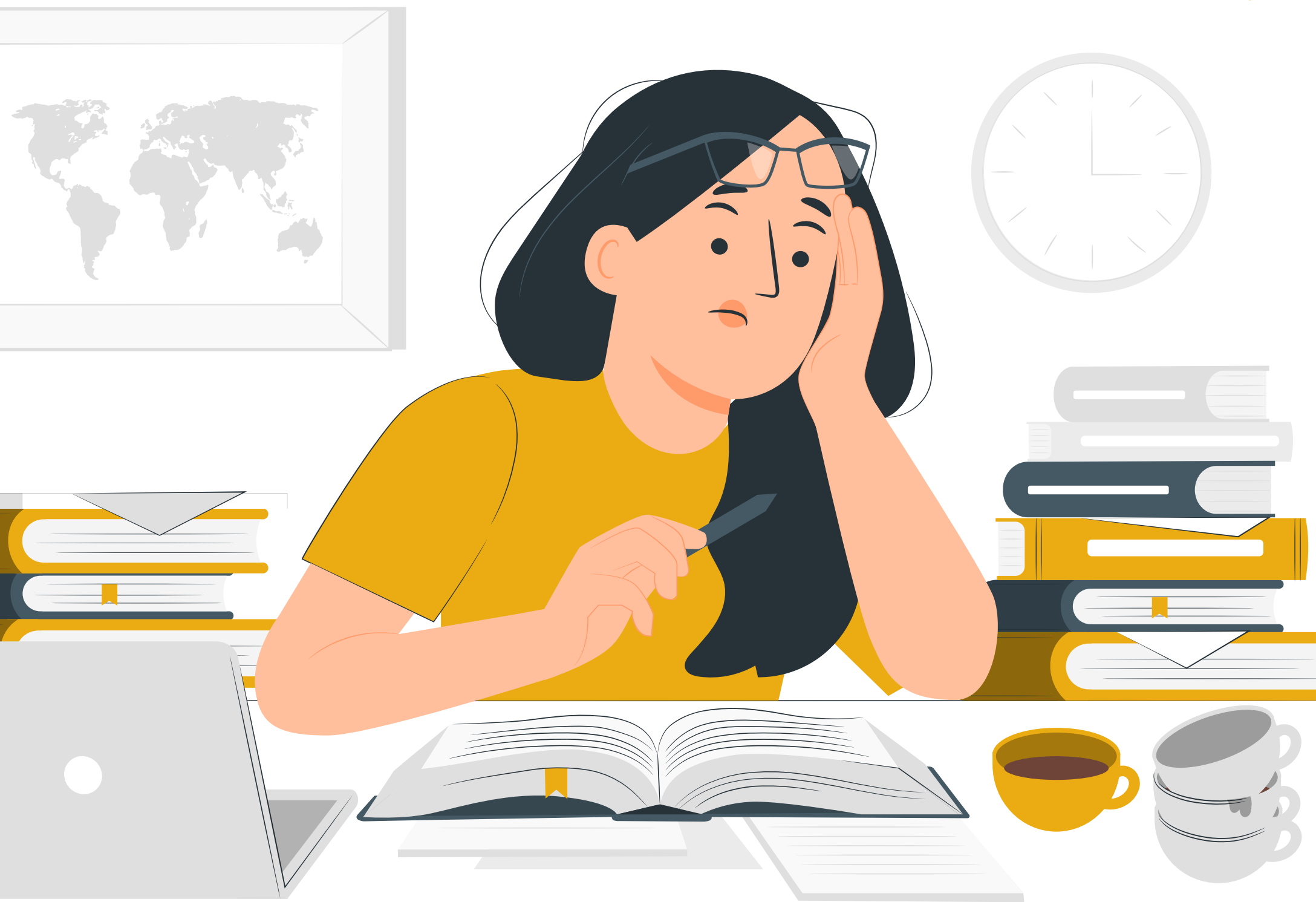
- Does the application integrate evidence-based practices supported by current research (e.g., CBT, DBT, ACT)?
- Are peer-reviewed publications available?
- How was effectiveness measured? Was the sample representative of the student population? Were outcomes reported for diverse demographic groups?
- Are metrics tracked beyond the reduction of negative experiences, such as improvements in hope, agency, self-efficacy, or well-being?

User agency

- Does the tool promote autonomy, not dependence?
- Are students encouraged to build skills and engage in real-world behavior change?
- Does it avoid prescriptive advice and help users arrive at their own insights?
- Does it promote social interaction, integration, and belonging beyond the app?

Cultural inclusivity & equity

- Are demographic-based biases (e.g., racism, ableism, homophobia) monitored?
- Was user testing conducted with a representative, diverse student sample?
- Are interactions designed with cultural sensitivity?
- Are outcomes tracked by demographic subgroup?



Conclusion

With rising student demand, shrinking resources, and increasing scrutiny, higher education leaders are seeking scalable, ethical solutions to support student mental wellness. Students are already turning to AI — often before institutions are ready — making thoughtful tool selection more urgent than ever.

But not all AI mental wellness tools are created equal. Some fall short on safety, inclusivity, or measurable outcomes, even when marketed as “fit-for-purpose.” Without clear criteria, institutions risk adopting tools that do not align with their mission or meet the needs of today’s students.

When intentionally chosen, GenAI tools can expand access, promote real-world behavior change, and strengthen student engagement, belonging, and retention. This guide is designed to help institutions evaluate those tools with rigor, transparency, and student well-being at the center — not just to adopt new technologies, but to do so responsibly.

References

Abelson, S., Eisenberg, D., & Johnston, A. (2024). *Digital mental health interventions at colleges and universities: understanding the need, assessing the evidence, and identifying steps forward. The Hope Center, Temple University. May 2024.*

Golden, A., & Aboujaoude, E. (2024). Describing the Framework for AI Tool Assessment in Mental Health and Applying It to a Generative AI Obsessive-Compulsive Disorder Platform: Tutorial. *JMIR Formative Research, 8*(1), e62963.

Reyes-Portillo, J.A., So, A., McAlister, K., Nicodemus, C., Golden, A., Jacobson, C., & Hubert, J. (In press.) Generative artificial intelligence powered mental wellness chatbot for college student mental wellness: An open trial. *JMIR Formative Research.*

Rousmaniere, T., Li, X., Zhang, Y., & Shah, S. (2025, March). *Large language models as mental health resources: Patterns of use in the united states.*

Scholich, T., Barr, M. M., Stirman, S. W., & Raj, S. (2024). Can Chatbots Offer What Therapists Do? A mixed methods comparison between responses from therapists and LLM-based chatbots.

Stade, E. C., Eichstaedt, J. C., Kim, J. P., & Stirman, S. W. (2025). Readiness Evaluation for AI-Mental Health Deployment and Implementation (READI): A review and proposed framework.