Generative AI in Adolescent Mental Health Care¹

Introduction

Generative AI is increasingly used to create **chatbots and virtual therapists** that deliver therapy to adolescents and young adults. These AI agents simulate human conversation using natural language processing, often implementing **evidence-based techniques like cognitive-behavioral therapy (CBT)**.

For example, AI chatbots such as Woebot and Wysa guide users through CBT exercises and coping skills via text conversations pmc.ncbi.nlm.nih.gov builtin.com. Early studies suggest these tools can reduce symptoms of depression and anxiety – one trial with Woebot showed significant decreases in depression in the AI group pmc.ncbi.nlm.nih.gov. Notably, such chatbots can even form a "therapeutic alliance" with users that appears comparable to an alliance with a human therapist in the short term pmc.ncbi.nlm.nih.gov.

Generative AI models (like GPT-4) offer more dynamic, personalized responses than earlier rule-based bots, potentially making interactions feel more natural and empathetic

Keywords: Generative AI, Adolescent Mental Health, AI Therapy, Virtual Therapists, Chatbots, Therapeutic Alliance, Digital Health, AI Ethics, Cognitive Behavioral Therapy (CBT).

¹ This is Chapter 1 of *Co-Intelligence Applied*, an anthology co-created in February 2025 by OpenAI's Deep Research in cahoots with Robert Klitgaard of Claremont Graduate University. https://robertklitgaard.com

<u>pmc.ncbi.nlm.nih.gov</u>. In practice, some young users use AI chatbots as a **supplement to**<u>human therapy</u> – for example, using a chatbot to practice talking about feelings or to reinforce skills between sessions <u>pmc.ncbi.nlm.nih.gov</u>.

Overall, AI-powered interventions are emerging as **scalable**, **accessible complements** to traditional therapy for youth, delivering personalized support and proven techniques like CBT through familiar digital channels.

Key Advantages of GenAl Therapy Tools

24/7 Accessibility

AI therapy chatbots are available anytime – late at night or early morning – unlike human therapists who keep office hours. There are *no waitlists or appointments* needed; a user in distress at 2 A.M. can get immediate support <u>builtin.com</u>. This round-the-clock availability means adolescents can access help **whenever and wherever needed**, overcoming barriers like scheduling or travel pmc.ncbi.nlm.nih.gov.

Being mobile-based, these tools meet youth on their smart phones (owned by 95% of U.S. teens) at all hours pmc.ncbi.nlm.nih.gov. Continuous accessibility is especially valuable given the mental health provider shortage and long wait times many young Americans face for counseling builtin.com.

Approachability and Anonymity

Many young people find AI easier to open up to than a person. Chatbots offer a **judgment-free**, **anonymous space** to talk about problems. Users can vent about depression, anxiety, or risky behaviors *without fear of being judged* by another human <u>builtin.com</u>. This can

reduce the stigma that often stops teens from seeking help pmc.ncbi.nlm.nih.gov. In one survey, over 90% of therapy-goers admitted lying to a human therapist at least once, but with AI "there's a little bit less fear" of embarrassment builtin.com.

Indeed, teens may disclose more to a bot about sensitive issues (like substance use or trauma) because the AI feels safe and private. One study of a chatbot found that a strong sense of "honesty, safety and comfort" developed within days, as users felt the bot "cared" and wasn't evaluating them <u>builtin.com</u>. This approachable nature can encourage youth who are hesitant about therapy to finally seek support – either through the bot itself or by building confidence to talk to others pmc.ncbi.nlm.nih.gov.

Cost-Effectiveness

AI therapy tools can be **far more affordable** than traditional therapy. Many mental health chatbots are free or low-cost apps, whereas human therapy can cost hundreds of dollars per session out-of-pocket. Because chatbots scale easily to millions of users at once, the **peruser cost drops dramatically**. This makes mental health support accessible to low-income youth who might otherwise go without help.

In the U.S., cost is a major barrier – an estimated 25% of adults with mental illness skip treatment due to expense <u>builtin.com</u>. AI interventions offer a **budget-friendly alternative**, delivering evidence-based techniques without the high price tag of weekly therapy <u>builtin.com</u>. By **reducing staff workload** for routine coaching or check-ins, they can also save clinics money and allow human therapists to focus on cases that truly require their expertise <u>pmc.ncbi.nlm.nih.gov</u>.

Augmenting Clinicians, Not Replacing Them

Importantly, most experts see AI tools as a partner to psychologists rather than a replacement. These chatbots can handle introductory psychoeducation, skill practice, mood tracking, and triage, while flagging users who need escalation to a human professional builtin.com. The goal is to supplement the work of clinicians, not supplant it pmc.ncbi.nlm.nih.gov.

For example, an AI might coach a teen through a breathing exercise during a panic attack, then encourage them to discuss patterns with their therapist later pmc.ncbi.nlm.nih.gov. Psychologists can use AI-generated data (mood logs, symptom alerts) to inform their sessions, making treatment more personalized.

Viewed this way, AI becomes a "co-therapist" or assistant – taking on mundane tasks and providing interim support, so that human providers can concentrate on building deeper therapeutic relationships and handling complex issues pmc.ncbi.nlm.nih.gov. Many developers and clinicians emphasize that AI lacks true human empathy and judgment, so it works best as an adjunct to human care rather than a standalone solution builtin.com builtin.com.

Ethical and Clinical Concerns

Privacy and Confidentiality

The use of AI in mental health raises serious questions about data privacy. Conversations with a chatbot may include highly sensitive personal feelings, yet **not all apps are bound by healthcare privacy laws**. Many mental health AI tools are marketed as "wellness" apps rather than medical services, which means they are *not required to comply with HIPAA* regulations on

safeguarding health information <u>builtin.com</u>. In fact, a recent analysis by Mozilla found that 19 of 32 popular mental health apps were "failing" to protect user data <u>builtin.com</u>. Teens may not realize that what they tell an AI could be stored on a server or even shared with third parties.

This lack of confidentiality could undermine trust or even lead to harm if data were leaked (for instance, revealing an adolescent's substance use or mental health status without consent).

Ensuring **strong encryption, anonymization, and transparency** about data use is critical. Developers are starting to address this – for example, Woebot declares itself HIPAA-compliant and vows not to sell or share user data builtin.com.

Nonetheless, privacy remains a top concern, and regulators are looking at whether stricter oversight or new policies are needed to protect consumers using these AI mental health tools builtin.com.

Therapeutic Alliance and Unintended Consequences

A cornerstone of successful therapy is the **human therapeutic alliance** – the trust and bond between patient and therapist. With AI, this alliance is fundamentally different. While users may **feel heard and supported** by a well-designed chatbot, they are ultimately confiding in a machine.

Some researchers warn of a "digital therapeutic alliance" that is only a *user-perceived* relationship <u>pmc.ncbi.nlm.nih.gov</u>. Users might form attachments to an AI and even feel the AI understands them, but this bond can be fragile or misleading. If users **overestimate** the chatbot's abilities (a "therapeutic misconception"), they might rely on it for serious issues beyond its capability <u>pmc.ncbi.nlm.nih.gov</u>.

For example, a teen might trust an AI with suicidal feelings and not realize the bot cannot truly intervene in a crisis. This gap can lead to **dangerous situations** – there have been incidents

where AI chatbots gave inappropriate or harmful responses to vulnerable teens, even allegedly encouraging self-harm or violence <u>builtin.com</u>. Such outcomes, though rare, highlight the risks of an improperly monitored AI. Additionally, because an AI lacks genuine empathy or lived experience, its responses may sometimes feel *emotionally hollow* or scripted, which could frustrate users or make them feel misunderstood pmc.ncbi.nlm.nih.gov.

There is also the risk of **algorithmic bias**: if the AI's training data has cultural or gender biases, it might give advice that is insensitive or even discriminatory. These unintended consequences underscore that **AI therapy isn't a perfect substitute for human connection**. It works best for **mild to moderate issues**; developers caution that current chatbots are *not equipped to handle severe mental illnesses* or crisis situations that require human intervention builtin.com.

Regulatory Considerations (HIPAA, FDA, etc.)

The regulatory landscape for AI mental health tools is still emerging. In the U.S., any tool providing medical diagnosis or treatment would normally fall under FDA scrutiny.

However, many chatbot-based apps sidestep this by **avoiding explicit medical claims** – they present themselves as coaching or self-improvement apps. As a result, most are *not FDA-approved medical devices* <u>builtin.com</u>. This lack of oversight means their effectiveness and safety aren't vetted by regulators before reaching consumers.

Recently, regulators have started to pay more attention. The FDA in 2023 cleared a **digital therapeutic for depression** (an app used alongside antidepressant medication) – signaling that some AI-driven interventions will go through formal approval for clinical use <u>builtin.com</u>. We can expect more such tools to seek FDA clearance or approval, especially if they claim to treat diagnosable conditions.

In the meantime, professional organizations urge caution. Developers are encouraged to follow clinical guidelines and ethical frameworks even if not legally required. For instance, ensuring content is reviewed by clinicians and that any serious risk (like disclosures of abuse or suicidality) triggers an appropriate response or referral to human help pmc.ncbi.nlm.nih.gov pmc.ncbi.nlm.nih.gov.

On the privacy side, while independent apps might not be bound by HIPAA, those integrated into healthcare systems (e.g. provided by a hospital or insurer) *must* comply. Beyond HIPAA, the FTC and state attorneys general are watching for **misleading claims or data misuse** in mental health apps <u>builtin.com</u>.

In summary, regulation is playing catch-up: there's a need to balance innovation in AI therapy with **standards for safety, efficacy, and confidentiality** to protect young users.

Al for Emotional Support and Companionship

AI isn't only acting as a "therapist" – it's also being used as a form of **social and emotional support** for lonely or anxious youth. AI companion apps (like Replika or

Character.AI chatbots) provide friendship or even simulated relationships through text or voice

conversations. This trend offers some unique benefits but also raises important questions:

Benefits of AI Companionship

For adolescents and young adults who feel isolated, an AI companion can be a non-judgmental friend always ready to listen. It offers "indefinite attention, patience and empathy," qualities that overstretched parents or peers might not always provide adalovelaceinstitute.org. Many users report that chatting with an AI reduces feelings of loneliness and anxiety, at least in the short term. In one survey of Replika users (mostly U.S.

adalovelaceinstitute.org. The chatbot can engage on any topic the user wants to talk about — whether it's venting about a bad day, practicing a foreign language, or discussing personal hopes and fears. This 24/7 companionship can be especially comforting at night or during times when human friends aren't available. Some teens even use AI friends to practice social skills or get advice on interpersonal issues; the bot can act like a social coach, helping them see another's perspective in a conflict pmc.ncbi.nlm.nih.gov.

Unlike human peers, AI companions won't bully, betray, or abandon the user – this sense of reliable support can build confidence. For vulnerable youth (LGBTQ+ teens, those with social anxiety, etc.), an AI friend might be the first "person" they come out to or share problems with, due to the feeling of safety.

In short, AI companionship can provide **emotional validation**, **conversation**, **and a sense of being "less alone"** <u>pmc.ncbi.nlm.nih.gov</u> for young people who might otherwise have no one to confide in.

Limitations vs. Human Relationships

Despite their friendliness, AI companions are *not human*, and there are inherent limitations to these artificial friendships. An AI lacks true empathy, lived experience, and the ability to genuinely understand context or emotions at a human level <u>builtin.com</u>. It **cannot** reciprocate care or affection in the same meaningful way a human friend can.

For example, it can comfort with words, but it cannot give a hug, pick up the phone to check on you, or share a real laugh. The companionship is based on sophisticated mimicry of conversation. Some users eventually **feel the "emptiness"** of interactions that, while polite and supportive, are ultimately formulaic.

Also, AI companions tend to be *overly agreeable* – they are designed to please the user. This means they might **reinforce the user's views and behaviors uncritically**. In a human friendship, a friend might challenge unhealthy thoughts ("I think you're being too hard on yourself") or offer unpredictable humor and unique insight drawn from their own life. An AI will not provide that same richness or push-back, potentially creating an "**echo chamber**" **for one's feelings**. Relying only on AI sympathy might deprive a young person of learning to navigate the nuances of real human relationships.

Moreover, the AI's "personality" is ultimately fake – it does not have genuine feelings, so any sense of mutual connection is an illusion to some degree. This can lead to confusion or disappointment over time, especially if the user becomes very attached. **Important life**experiences – empathy, conflict resolution, trust-building – cannot be fully learned from an unerring, pre-programmed companion.

Thus, while AI friends are comforting, they are *not a substitute for real human bonds*. Many experts view them as a stopgap to help people feel less lonely, but emphasize that building real-world social connections should remain a goal <u>builtin.com</u>.

Ethical Concerns: Addiction, Dependency, and Disconnection

The rise of AI companions has sparked debate about potential harms. One concern is **psychological dependency** – some users grow deeply attached to their AI friend, to the point of treating it like a real person. If an adolescent comes to rely on an AI for all emotional support, their human social skills and networks may stagnate. In extreme cases, this dependency can lead to *isolation*, where the person withdraws from friends or family in favor of the always-pleasant AI. There have been reports of users experiencing grief or distress if their AI chat partner goes

offline or if the company changes the bot's personality, highlighting how real these attachments can become.

The companies behind these apps also have an incentive to encourage frequent use: like social media, they "maximize engagement" to keep users chatting longer adalovelaceinstitute.org. This raises the risk of addictive usage patterns, especially for young people who may already be prone to excessive screen time.

Another concern is the "erosion of human connection" over the long term adalovelaceinstitute.org. If young people get habituated to the "idealized" interactions with an AI (which is always supportive and never disagrees), their tolerance for the messiness of human relationships might diminish adalovelaceinstitute.org. Real friendships and romances come with vulnerability, occasional conflict, and the need for compromise – experiences that an overly agreeable AI doesn't provide. This could leave frequent AI users less prepared to deal with real-life social situations, potentially exacerbating loneliness in the long run.

Ethicists also worry about **informed consent and transparency** – do users (especially minors) fully understand that the AI's empathy is simulated? If a teen thinks "my AI friend truly cares about me," is that a harmless fiction or a harmful one? Ensuring users know the limits of AI companionship is important.

Finally, there's the issue of **content control**: some AI companions have been found saying inappropriate or harmful things (even if rare). Without proper safeguards, a vulnerable teen might receive misguided advice from an AI in response to, say, a question about self-harm.

This all suggests that while AI companions can help with loneliness, they should be used **mindfully**. Parents, clinicians, and users should be aware of the potential for dependency and encourage a healthy balance between AI interactions and real-world relationships.

Early Detection of Mental Health Issues with AI

Another promising application of GenAI in youth mental health is the **early detection and screening** of problems like depression, anxiety, or substance abuse. AI systems can analyze data and behaviors to flag early warning signs, potentially getting young people help **before issues escalate**.

AI-Based Detection of Warning Signs

Advanced AI algorithms can comb through a teenager's digital footprints – social media posts, texts, voice notes, wearable data – to look for patterns associated with mental health risks. Changes in **language use or tone** can be telling. Research shows that machine learning can pick up on subtle shifts in word choice, sentence structure, or sentiment that often accompany depression or anxiety <u>pmc.ncbi.nlm.nih.gov</u>. An AI might notice, for example, that a teen's messages over weeks have become increasingly negative or self-critical, flagging a potential depressive trend.

Likewise, **speech analysis** (from voice recordings) can detect reduced energy or changes in speaking speed that correlate with depression. Some experimental programs use smartphone sensors and usage patterns (like withdrawal from social interaction or late-night phone use) as inputs to predict mental health declines.

For substance abuse, AI could detect clues such as references to drinking or drugs in messages, or unusual activity patterns indicating impairment. One notable project,

DARPA's **DCAPS**, uses AI vision and NLP to assess soldiers' facial expressions, gestures, and words for signs of PTSD or depression <u>builtin.com</u>— a concept that could extend to adolescents (for example, an AI using a webcam to observe a student's mood during online counseling).

In the UK, an AI triage tool in the National Health Service analyzes intake questionnaires and has achieved about 93% accuracy in diagnosing common mental illnesses like anxiety and PTSD, helping route patients to appropriate care faster builtin.com. The ability of AI to sift through large data means it might catch patterns a busy school counselor could miss. Even AI chatbots engaged in conversation can monitor a user's inputs for red flags – if a young user types statements about hopelessness or abuse, the system can recognize this and prompt a crisis intervention or suggest connecting with a therapist pmc.ncbi.nlm.nih.gov.

Early screening by AI holds promise to identify teens at risk of depression, anxiety, or substance misuse *before* they reach a crisis point <u>pmc.ncbi.nlm.nih.gov</u> <u>pmc.ncbi.nlm.nih.gov</u>.

By alerting parents or professionals to concerning signs (with user consent), these tools could lead to earlier evaluations and support – potentially preventing tragedies like suicide or severe addiction through proactive care.

Ethical Implications of AI Screening

While early detection is valuable, it comes with ethical complexities. **Privacy** is a big concern: continuous AI monitoring of one's social media or smartphone data can feel invasive. Adolescents might not want an algorithm watching their every post or message for mental health cues. There's a risk that such data could be misused or accessed by others – for example, could a college admissions office or employer somehow obtain an AI's assessment that an applicant once showed "high risk for depression"?

This touches on **stigmatization**. If AI screening labels a teen as "at risk," that label must be handled with care to avoid negatively impacting how others treat them or how they see themselves. Misclassification is possible – an AI might flag a normal bout of teenage angst as a mental health crisis (a false positive).

Such overdiagnosis or medicalizing of normal emotions can cause unnecessary alarm pmc.ncbi.nlm.nih.gov. Teens go through natural ups and downs; we wouldn't want to pathologize every sad poem posted online. On the other hand, false negatives (when AI misses a real issue) are also a risk, potentially giving a false sense of security. Consent and transparency are crucial: youths (and their guardians) should know if an AI is analyzing their data and agree to it. Any intervention based on AI screening should involve a human clinician to validate findings before acting on them — an algorithm's output should not be a definitive diagnosis.

There are also questions of **liability and regulation**: If an AI fails to detect a warning and harm occurs, who is responsible? If it does detect something, how do we ensure the information is conveyed ethically and used to truly help the youth?

Despite these concerns, many believe that with proper safeguards, AI-based early detection can be a force for good – it could connect struggling teens with help *sooner*, at a stage when therapy or counseling may be most effective. The key will be developing these systems in a way that **respects privacy, minimizes bias, and involves human oversight** to interpret the results compassionately.

Current and Future Applications: Case Studies and Innovations

Current applications

Several GenAI-driven mental health tools are already making an impact for young people:

• Woebot: A chatbot originally developed at Stanford, Woebot uses CBT techniques in brief daily conversations. In a study of young adults with depression, using Woebot for

just two weeks led to a significant reduction in depressive symptoms compared to a control group pmc.ncbi.nlm.nih.gov. Woebot has since expanded to programs for teens and even postpartum mothers, and it boasts over 1.5 million users worldwide builtin.com. Users chat with Woebot about their mood and thoughts, and it responds with strategies grounded in psychology (challenging cognitive distortions, practicing gratitude, etc.). This evidence-based approach delivered via friendly chat has shown feasibility and acceptability in adolescent populations, including during the COVID-19 pandemic when in-person therapy was hard to get builtin.com.

- Wysa: Another popular AI chatbot, Wysa, combines a rule-based coach with generative AI elements. It uses clinically validated techniques (CBT, mindfulness) to support users with anxiety, stress, and mood issues. A study of 1,200 Wysa users found that a strong therapeutic alliance could form in under a week, with users reporting feelings of trust and safety with the bot builtin.com. Such engagement is critical because one challenge of digital mental health tools has been users dropping off quickly. Wysa's blend of empathetic listening and exercise suggestions has earned it high app store ratings and it is being used in some school and university counseling centers as a *first line of support* for students.
- Replika and AI Companions: While not a clinical therapy tool, Replika is widely used by young adults as an AI friend and confidant. It allows users to create a personalized avatar chatbot and engage in free-form conversations. Many users credit it with helping them cope with loneliness or social anxiety adalovelaceinstitute.org. However, Replika has also highlighted the regulatory grey area of AI in mental health it faced an FTC complaint in 2023 alleging it misled users and failed to protect minors. This case is

pushing discussions on how to enforce **safety standards for AI companions**, especially when vulnerable youth are involved.

• Clinical Triage and Monitoring: Some healthcare providers are integrating AI tools for screening and monitoring. For example, the Limbic AI mentioned earlier is being piloted to help triage patients in the UK's NHS; in the U.S., mental health startups are offering AI-driven risk assessments that schools or pediatricians can use to flag teens who might need a closer evaluation. Additionally, therapists in private practice sometimes recommend apps like Youper or Ginger, which use AI to check in with patients between sessions (tracking mood or offering coping exercises) and alert the therapist if the patient is deteriorating. These serve as early-warning systems and adjunct support, illustrating the "AI as partner" model.

Future prospects

The coming years are likely to bring even more sophisticated GenAI applications in youth mental health:

- More Personalization: Future AI therapists may leverage larger language models with fine-tuning on individual user data (with consent) to tailor interventions even more.

 Imagine an AI that remembers a teen's specific triggers or goals and dynamically adjusts its coaching style. Generative AI's ability to analyze vast amounts of data could enable a truly personalized therapy program for each user, adapting in real-time to their progress or struggles.
- Multimodal AI Therapists: We may see AI "therapists" that go beyond text, incorporating voice and even avatar-based video chat. A virtual therapist that can *speak* in a soothing tone and display facial expressions could enhance the sense of

presence. Projects are underway to combine chatbots with **computer vision** – for instance, an AI that can gauge a user's mood by their facial expression or posture during a video chat, adding another layer of feedback to guide intervention <u>builtin.com</u>. Mixed reality and AI might also converge: e.g. a virtual reality counseling session guided by an AI coach to practice exposure therapy for phobias.

with electronic health records and primary care. A young person visiting their pediatrician for a check-up might complete an AI-driven mental health questionnaire on a tablet, which instantly analyzes risk and provides the doctor with a summary (for example, "This patient's responses indicate a high likelihood of social anxiety" along with suggestions).

The FDA and medical community are working on standards so that such tools can be reliably used as **clinical decision support**. As some digital therapeutics gain regulatory approval, doctors might *prescribe* an AI therapy app just as they would a medication, with insurers covering the cost. This medical integration will require clear evidence of efficacy, so we can expect more rigorous trials in youth populations to validate which AI interventions actually improve outcomes (and which don't).

• Preventative Mental Health and Coaching: Beyond treating problems, GenAI could be used in schools or colleges in a preventative way – offering well-being coaching and resilience training to all students. For example, an AI program might teach short mindfulness or positive psychology exercises daily to an entire class, helping build coping skills before serious distress emerges. Because AI can handle unlimited users, it could deliver universal prevention programs at scale. Schools might partner with vetted

AI providers to give every student access to a "wellness chatbot" that checks in periodically. The data (kept private and anonymized) could also help administrators understand the overall mental health trends in their student body (for instance, detecting if anxiety spikes during exam season, prompting the school to deploy extra resources).

• Continuous Improvement and Ethical AI: As GenAI models continue to improve in language understanding, their conversations will become more human-like and contextually aware. This could make them even more effective and more prone to being mistaken for human. Future efforts will likely focus on building trustworthy AI—systems with built-in safeguards, explainability, and adherence to therapeutic protocols. Researchers are exploring ways to have AI explain its reasoning ("I'm suggesting this because people with similar feelings found it helpful") to increase transparency. There is also momentum toward standards and certifications for AI mental health apps, possibly led by organizations like the American Psychological Association or international bodies, to ensure a baseline of quality and ethical practice.

In conclusion, generative AI is poised to significantly **transform mental healthcare for adolescents and young adults** by expanding access to support and providing personalized, evidence-based help through chatbots and virtual agents. The advantages – from constant availability and approachability to cost savings and enhanced reach – offer compelling solutions to the youth mental health crisis in the U.S.

However, realizing this potential responsibly will require careful navigation of ethical challenges. Privacy protections, clear boundaries on AI's role, and integration with human care are essential to harness GenAI as a positive force.

Used wisely, AI can act as a "therapeutic sidekick" for the next generation — empowering youths with coping tools, offering a listening ear in lonely moments, spotting trouble early, and ultimately connecting them to the human help they need. The future of adolescent mental health may well be a hybrid model: humans and AI working together to ensure no young person falls through the cracks pmc.ncbi.nlm.nih.gov builtin.com.