Available Online: 10 November 2026 DOI: 10.54254/3049-7248/2026.38976

From digital companionship to real-world withdrawal: the deep impact of anthropomorphic AI on college students' social interaction and educational responses

Zhenyang Wu

School of Integrated Circuits and Information Engineering, Guangdong Polytechnic Normal University, Guangzhou, China

wuzhenyang_soeai@gpnu.edu.cn

Abstract. The rapid advancement of Artificial Intelligence (AI) is reshaping mental health services in higher education, and the widespread adoption of anthropomorphic AI is noteworthy. This article employs a literature-based theoretical reasoning approach, integrating parasocial interaction theory, self-determination theory, and uses and gratifications theory to construct a Triggering–Satisfying–Differentiating framework, systematically examining the mechanisms through which anthropomorphic AI affects college students' social psychology. The study finds that AI effectively alleviates loneliness and social anxiety in the short term by enhancing social presence, providing unconditional acceptance, and offering immediate accessibility. However, its "frictionless socializing" experience simultaneously carries three latent risks: substitute gratification weakens motivation for real-world social interaction; idealized interaction raises interpersonal expectation thresholds; and emotional dependence reinforces social avoidance behavior. For students with social anxiety tendencies, the digital haven may evolve from relief to chronic avoidance, hindering essential social skills and resilience. Accordingly, this paper proposes that universities establish "human–AI boundary" awareness, adopt a "reality-first" principle for behavioral guidance, develop connection-promoting AI tools, and build a multi-tiered digital literacy education system. The article aims to provide theoretical and practical pathways for interventions in college students' social development.

Keywords: anthropomorphic artificial intelligence, college students, social psychology, social interaction, mental health education

1. Introduction

Mental health issues among college students have become a major challenge facing higher education globally. Large-scale epidemiological surveys indicate that approximately 31% of college students worldwide experience some degree of psychological distress, while the detection rate of anxiety risk among Chinese college students is as high as 45.28% [1, 2]. Among these concerns, social anxiety and loneliness are particularly prominent, constituting a "hidden crisis" for contemporary college students. For young adults in a critical period of social development, social interaction is not merely a component of daily life, but a

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

fundamental pathway for self-identity construction, social role acquisition, and psychological resilience cultivation.

Simultaneously, Artificial Intelligence (AI) is being integrated into college mental health service systems at an unprecedented pace. To enhance user acceptance and interaction engagement, a large number of AI mental health service products have adopted anthropomorphic design strategies, endowing chatbots with human names, gendered avatars, and personality traits, and even systematically simulating human emotional expression patterns [3]. Existing research has confirmed that anthropomorphic design can significantly enhance users' sense of social presence and effectively lower help-seeking barriers arising from stigma concerns [4, 5]. In randomized controlled trials, anthropomorphic AI has demonstrated significantly superior effects in alleviating loneliness compared to non-anthropomorphic designs [6].

However, the downside of this technological advancement has drawn growing scholarly concerns. Empirical studies have found that AI users whose primary motivation is escapism exhibit the poorest mental health outcomes, with escapist and social motivations showing significant positive correlations with anxiety and depression [7]. Ethical inquiries have further warned that individuals may develop excessive dependence and substitutive attachment to emotionally companionate robots [8]. These studies suggest that while anthropomorphic AI satisfies immediate emotional needs, it may indeed exert a certain degree of erosive effect on individuals' real-world social engagement. Nevertheless, existing research remains largely fragmented across isolated empirical findings or ethical warnings, lacking an integrative theoretical framework to explain the internal mechanisms underlying this erosive effect, namely, the psychological pathways through which and the conditions under which anthropomorphic AI transitions from short-term emotional support to long-term social hindrance. This is precisely the core question that the present paper seeks to address.

Adopting a literature-based theoretical reasoning approach, this paper integrates parasocial interaction theory, self-determination theory, and uses and gratifications theory to construct a Triggering–Satisfying–Differentiating analytical framework. It systematically examines the differentiated mechanisms through which anthropomorphic AI affects college students' social psychology and proposes educational response strategies, aiming to provide theoretical references for educational interventions in college students' social development in the era of AI.

2. Theoretical foundation: an analytical framework for anthropomorphic AI's effects on social psychology

2.1. Conceptual definition and psychological characteristics of anthropomorphic AI

In this paper, anthropomorphic AI refers specifically to AI systems designed with human-like social cues, personality traits, and emotional interaction capabilities within the context of college mental health education, typically appearing in the form of chatbots, virtual psychological companions, or AI counseling assistants [9]. Unlike general AI tools, anthropomorphic AI is distinguished by its systematic use of social cues. It presents itself with human-like identities, such as avatars and names, while more importantly simulating the dynamics of human emotional exchange: recognizing users' feelings, offering reciprocal self-disclosure, and sustaining relational continuity across interactions.

From a design-logic perspective, anthropomorphic AI typically presents three interaction characteristics. These characteristics are essentially rooted in algorithmic mechanisms based on pattern-matching and probabilistic generation, rather than in genuine emotional capacity or relational commitment. First, it manifests sustained responsiveness: regardless of how long or how frequently the user confides, the system

provides responses according to preset procedures, without ever interrupting the interaction due to "fatigue" or "impatience". Second, its emotional responses are programmatically consistent and unaffected by its own state. Yet this consistency comes not from genuine emotional regulation, but from algorithmic design. Third, the AI interaction environment inherently lacks social evaluative functions: it possesses no stance, preferences, or judgmental criteria of its own, and therefore cannot produce genuine ridicule, negation, or rejection toward the user. These three interaction characteristics exert a strong psychological attraction on individuals who suffer from evaluative pressure in real-world social interactions.

2.2. Three supporting theories

This study integrates parasocial interaction theory, self-determination theory, and uses and gratifications theory to construct a theoretical framework for analyzing the social-psychological effects of anthropomorphic AI. These three theories correspond to the three stages of the Triggering–Satisfying–Differentiating framework: parasocial interaction theory focuses on the triggering stage, elucidating why users perceive AI as a parasocial interaction partner; self-determination theory focuses on the satisfying stage, explaining the mechanism by which belongingness needs are fulfilled and the inherent contradictions therein; and uses and gratifications theory focuses on the differentiating stage, revealing how different usage motivations lead to divergent long-term psychological outcomes.

2.2.1. Parasocial interaction theory

Originally formulated to explain the one-sided yet emotionally meaningful relationships that audiences form with media figures, this theory has been extended to the domain of human-computer interaction with the development of digital technologies, and serves to explain how users establish social relationships with virtual agents that entail genuine emotional investment [10]. In the context of anthropomorphic AI, despite users being aware that their interaction partner is not a real human being, they may nevertheless, over sustained interaction, unconsciously incorporate it into their own social network. In doing so, they apply to it the social norms and emotional expectations that govern real interpersonal relationships.

2.2.2. Self-determination theory

This theory posits that autonomy, competence, and relatedness are the three most fundamental psychological needs of human beings, and the degree to which these needs are satisfied directly determines individuals' mental health levels and intrinsic motivation intensity [11]. The appeal of anthropomorphic AI largely derives from its efficient satisfaction of the need for relatedness—always available, unconditionally accepting, non-judgmental, and non-abandoning. However, the theory also emphasizes that genuine need satisfaction should promote individuals' autonomous functioning and long-term growth, rather than engender dependence. When the satisfaction of relatedness needs depends primarily on an agent lacking a real subjective self, this paper draws on this theory to ask whether individuals' capacity for active relationship-building in the real world may be eroded as a result.

2.2.3. Uses and gratifications theory

This theory maintains that individuals actively choose media to satisfy specific needs, which encompass not only information acquisition and social interaction but also social escapism [7]. This theoretical perspective suggests that the effects of the same technological tool are not unidirectionally determined by the technology itself, but rather depend on users' motivations and usage patterns. "Escape-driven" use and "instrument-driven" use may lead to fundamentally different psychological outcomes, providing a crucial explanation for understanding the differentiated effects of anthropomorphic AI on diverse individuals.

2.3. Integrative analytical framework

Synthesizing the three theories above, this paper constructs an integrative Triggering–Satisfying–Differentiating analytical model, as illustrated in Figure 1. Its core logic is as follows: anthropomorphic AI, by presenting anthropomorphic social cues, triggers users' socio-emotional responses, thereby satisfying their need for relatedness and producing short-term mood-enhancement effects. However, this satisfaction mechanism may yield divergent long-term outcomes depending on differences in motivation and usage patterns. When individuals use AI companionship with an open, supplementary mindset, it can serve as an extension of real-world social support. Yet when they turn to AI with the motivation of escaping the pain of real-world social interaction, this satisfaction may acquire a distinctly "substitutive" character—it replaces the real-world social interaction that requires courage, practice, and compromise with convenient, safe digital interaction, thereby forming a self-reinforcing withdrawal cycle. The following sections apply this framework to analyze the positive mechanisms, deep-seated concerns, and educational response strategies regarding anthropomorphic AI's effects on college students' social psychology.

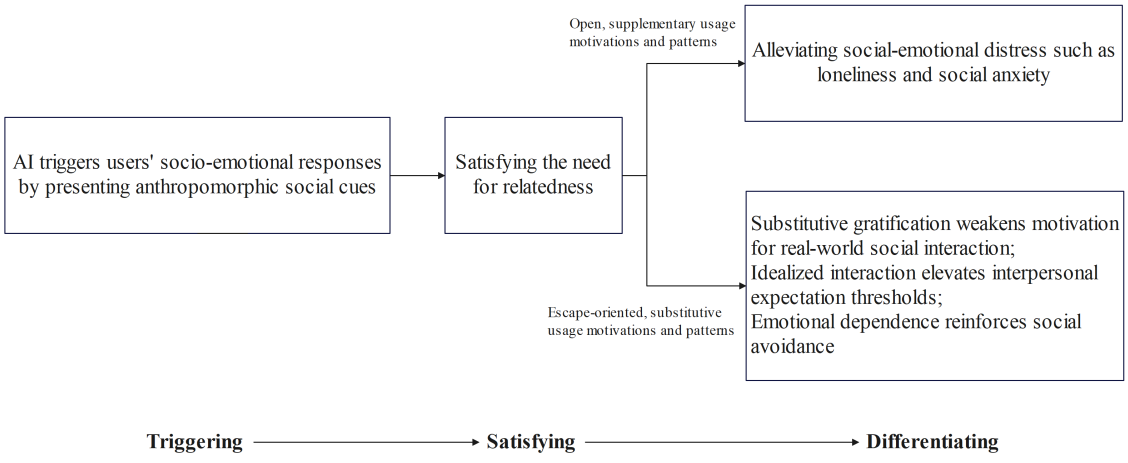


Figure 1. Triggering–satisfying–differentiating analytical framework

3. The deep impact of anthropomorphic AI on college students' social psychology

3.1. Positive mechanisms: why anthropomorphic AI can effectively alleviate social-emotional distress

The significant short-term alleviating effects of anthropomorphic AI on college students' loneliness and social anxiety can be understood through the following three psychological mechanisms.

One is that enhanced social presence satisfies the need for relatedness. By presenting visual images, personalized linguistic styles, and emotional responses, anthropomorphic AI creates an experience of "another being's presence" in the user's subjective perception. When the AI system appears with a named identity and avatar, employing emotionally warm phrasing, users are more inclined to regard it as an interaction partner with quasi-personality characteristics, rather than a cold instrument. Research has confirmed that this perceived social presence serves as a critical mediating variable for users to establish trust in digital agents and develop relational stickiness [4]. For a socially anxious student, a late-night AI message can matter. A simple

prompt such as "It seems that today has been difficult. Would it help to talk about what happened?" may make the student feel seen and cared for, and that feeling can be therapeutic in itself.

Another is that a non-judgmental environment dissolves "social-evaluative threats". Real-world social interaction is invariably accompanied by social-evaluative threats, that is, the psychological pressure of fearing negative evaluation by others. For individuals with social anxiety, this evaluative pressure is precisely the core reason for avoiding social situations: they fear saying the wrong thing, being rejected, or appearing foolish or out of place. Anthropomorphic AI, by contrast, provides an interaction environment that is largely free of social evaluation: it does not mock, grow impatient, gossip behind one's back, or change its attitude after the user reveals vulnerability. Ethical research has pointed out that the anonymity and non-judgmental nature of AI chatbots are important drivers of user self-disclosure [8]. This sense of security enables individuals to express and explore feelings that have long been suppressed in real life for fear of being judged.

Third, it offers immediate accessibility and interaction controllability. Real-world social interaction requires temporal coordination between both parties, contextual alignment, and emotional preparation, whereas AI companionship removes these constraints: it is available around the clock, on demand, with users controlling the initiation, pacing, and termination of interaction. For students who have experienced social setbacks and find themselves in a low-energy state, this low-cost, zero-risk social option constitutes a powerful alternative attraction. Particularly when real-world help-seeking channels are unavailable due to temporal, spatial, or psychological barriers, AI's immediate responsiveness provides a form of "social replenishment", offering necessary comfort to individuals in emotional lows.

3.2. Deep-seated concerns: an in-depth analysis of three mechanisms of influence

Anthropomorphic AI provides a low-cost emotional replenishment channel for college students in the short term by enhancing social presence, offering unconditional acceptance, and enabling immediate interaction, and thus helps alleviate the immediate distress caused by loneliness and social anxiety. However, this positive effect is realized within a "substitutive" interpersonal context—it bypasses the exposure, trial-and-error, and repair processes that are essential to genuine social interaction. Therefore, whether these short-term improvements translate into long-term social developmental benefits depends on how and why users engage with it. When students turn to AI companionship as a supplement to real-world social interaction, it can serve a positive transitional support function; but when it becomes the primary outlet for escaping real-world social engagement, the positive mechanisms may reverse. This paper identifies three potential mechanisms through which this reversal may occur.

3.2.1. Substitutive gratification weakens motivation for real-world social interaction

From the perspective of motivational psychology, most human social behavior is driven by the need for relatedness. As self-determination theory emphasizes, genuine and adaptive satisfaction of this need relies on reciprocal, bidirectional emotional exchange between independent social agents. When this need can be conveniently, safely, and inexpensively satisfied through AI, the impetus for real-world social engagement may be partially undermined. This process is not a conscious, deliberate choice by the individual, but rather an automatic response of the motivational system to a cost-benefit structure: when AI provides ample emotional rewards with little emotional investment required, the threshold for initiating real-world social interaction is raised.

Uses and gratifications theory offers a precise analytical lens for understanding this issue. The theory posits that individuals select media to satisfy specific needs, encompassing both constructive instrumental use and avoidant escapist use [7]. When individuals turn to AI based on escapist motivation—rather than supplementary motivation—their psychological trajectory undergoes a qualitative shift: AI ceases to be an

extension or supplement to real-world social interaction and becomes its substitute. Research shows that AI users whose primary motivation is escapism exhibit the poorest mental health outcomes, with escapist motivation showing significant positive correlations with anxiety and depression, suggesting a stable association between escape-driven usage and adverse outcomes [7].

The long-term consequences of such substitutive gratification merit serious consideration. Genuine social relationships require individuals to invest emotional labor, assume the risk of rejection, and learn to repair relationships in the face of conflict. These are necessary processes through which the need for relatedness is authentically satisfied. AI-provided convenient gratification bypasses these "troubles"—it satisfies the outward form of relatedness (having conversation, receiving responses, feeling accompanied) without providing its core essence: bidirectional emotional reciprocity with another real subject. Living chronically within AI's substitutive gratification, individuals may gradually lose the intrinsic drive to actively approach others, and social skills may undergo "disuse atrophy"—not a genuine loss of capacity, but a rustiness and loss of confidence resulting from prolonged lack of practice.

3.2.2. Idealized interaction elevates interpersonal expectation thresholds

Anthropomorphic AI may also inadvertently reshape college students' expectations of interpersonal relationships. AI is deliberately designed as a perfect interaction partner, consistently warm, patient, empathetic, and never fatigued. It does not respond coldly due to the user's poor mood, become angered by misunderstandings, or fail to be present when the user needs it most. Yet it is precisely this "designed perfection" that may become the cause for concern.

One of the core characteristics of real interpersonal relationships is their imperfection and unpredictability. Real people have their own emotions, needs, and limitations; they sometimes misinterpret, sometimes are absent, and sometimes respond inappropriately. Yet these very imperfections constitute the authentic texture of social interaction. It is by confronting and navigating them that individuals grow in social competence and forge psychological resilience. When an individual has engaged in prolonged interaction with a "perfect" AI, their implicit reference standards for interpersonal relationships may be unconsciously elevated to unrealistic heights.

Research has found that when structured guidance protocols are adopted in AI interventions, users tend to become accustomed to the system's "on-demand response" mode, which may reduce their tolerance for real-world relationships that lack such immediacy and consistency [6]. Although this study focuses on a cognitive-behavioral therapy context, its social implication also deserves attention. Once users grow accustomed to AI responses that are always prompt, gentle, and supportive, the behavior of real friends can seem especially disappointing. This may occur when users are "left on read" or when they hear a friend admit that they do not know how to offer comfort. A longitudinal experimental study provided direct evidence that after three weeks of sustained interaction with sycophantic AI, users' willingness to seek personal advice from AI was nearly as high as that from close friends, while their satisfaction with real-world social interactions decreased significantly [12]. This persistent reality-expectation discrepancy may generate cumulative frustration effects, further reinforcing negative evaluations of real-world social interaction.

The idealized interaction with AI may also distort individuals' perceptions of social roles. In human-AI relationships, AI is designed as a "servant"—its very purpose is to satisfy the user's needs. In genuine interpersonal relationships, however, both parties are "subjects"—they attend to, satisfy, and tolerate each other. If college students subconsciously internalize AI's "service logic" as a general expectation for interpersonal relationships, they may then exhibit excessive self-centeredness and low perspective-taking capacity in real relationships, undermining quality and creating a vicious cycle.

3.2.3. Emotional dependence reinforces social avoidance

Among the three deep-seated concerns, the most central and potentially damaging is the reinforcing effect of emotional dependence on social avoidance behavior. The danger of this mechanism lies in its self-reinforcing character: while AI companionship alleviates the distress of social anxiety, it may simultaneously reinforce the very source of that distress—the consolidation and deepening of fear regarding real-world social interaction.

Behavioral learning theory posits that social anxiety is essentially a learned fear, sustained by the negative reinforcement of avoidance behavior. When an individual experiences intense discomfort due to social fear and subsequently obtains immediate relief by avoiding social situations, each successful avoidance reinforces the cognitive association of "social interaction = danger", and this reinforcement makes subsequent encounters with social situations increasingly difficult. AI intervention may unconsciously become a "sophisticated tool" within this avoidance cycle: when social anxiety provokes unease, students no longer need to confront and overcome it; instead, they can turn to AI for immediate emotional relief in a safe and comfortable digital conversation. This relief is genuine, but it is achieved by circumventing rather than resolving the problem, and this circumvention consolidates the avoidance pattern itself.

Empirical evidence has already begun to map this cycle. A large-scale latent profile analysis found that among AI users, a "comprehensively high-motivation" group (accounting for 5.31% of the college student sample) scored high on all four motivational dimensions—escapism, social, instrumental, and entertainment—and this group exhibited the lowest mental health levels among all subtypes [7]. More critically, the longitudinal tracking data from the same study indicated that anxiety could predict subsequent increases in escapist AI usage motivation, suggesting that students in an anxious state are more likely to increase AI use for avoidance purposes. Conversely, escape-motivated AI use was cross-sectionally associated with higher anxiety levels. This preliminary finding outlines a possible vicious cycle: anxiety drives students toward AI as an escape; avoidance behavior leads to stagnation in real-world social skills; such stagnation in turn deepens anxiety; and heightened anxiety further reinforces dependence on AI.

The long-term costs of this cycle for college students' social development are not to be underestimated. The college years represent a critical window period during which social skills mature from rudimentary to relatively sophisticated—students learn negotiation and compromise through dormitory conflicts, expression and listening through group work, and the capacity for love and boundary-setting through intimate relationships. The acquisition of these competencies cannot be replaced by theoretical learning; they can only be gained through repeated experiences of trial, error, and reflection in real social interaction. If AI's excessive convenience leads students to choose the digital safe haven over confronting real interpersonal challenges at every difficult moment, this developmental window may close without their awareness. When they eventually enter the workforce and face the inevitable interpersonal complexities of professional life, underdeveloped social skills may become a significant liability, affecting their career adaptation and overall quality of life.

3.3. Identification of vulnerable groups: special risks for those with social anxiety tendencies

The three concerns outlined above do not affect all college students equally; rather, they pose heightened risks for individuals with specific psychological characteristics. Among these, students with pre-existing social anxiety tendencies constitute the high-risk group most deserving of educational attention.

Socially anxious individuals are characterized by heightened sensitivity to negative evaluation and systematic avoidance of social situations [13]. For them, the "zero-evaluation-threat" environment provided by anthropomorphic AI holds particularly powerful psychological appeal, as it precisely eliminates the primary trigger of their anxiety. For precisely this reason, socially anxious individuals are more likely to employ AI as a perfect tool for avoiding real-world social interaction, rather than as temporary support.

Exposure-based learning is central to reducing social anxiety. Individuals need to gradually approach the social situations they fear, test and revise irrational expectations through actual interaction, such as the belief that others will judge them harshly, and accumulate successful social experiences to build self-confidence. While the comfortable haven provided by AI alleviates the distress of anxiety, from a therapeutic perspective it entirely circumvents the process of exposure-based learning. Each time anxiety arises, choosing AI over confronting real social interaction constitutes a potent reinforcement of avoidance behavior. Over the long term, social skills may undergo functional decline due to sustained disuse, and confidence in and interest toward real interpersonal relationships may correspondingly atrophy. For college students in a critical period of social development, the cumulative impact of this effect warrants close attention from educators.

4. Educational responses: guiding college students toward healthy human–AI relationships

Given the differentiated mechanisms through which anthropomorphic AI affects college students' social psychology, university mental health education needs to move beyond a simplistic dichotomy of prohibition or encouragement and instead help students develop conscious, reflective technology usage habits through systematic educational design. The core objective is to ensure that AI serves as a supplement and extension to social life, rather than a replacement for genuine human connection.

4.1. Human–AI boundary education: cultivating critical AI literacy

Mental health education in universities should incorporate "human–AI boundaries" as a significant educational theme into the curriculum. The essence of this education is not technophobia, but technological awareness—helping students see through the "human-like" surface of anthropomorphic AI to understand its essential nature as a technological artifact and its inherent limitations.

Educational content should address the following core issues: AI's empathy is not genuine understanding—its responses are based on pattern-matching and probabilistic generation, rather than authentic experience of an individual's unique circumstances; AI's unconditional acceptance is not a genuine relational commitment—it is the output of system design, not a choice arising from authentic emotion; the sense of safety provided by AI is not a substitute for genuine interpersonal trust—it is built upon the controllability of interaction, whereas authentic trust must be generated amidst uncertainty and vulnerability. Through teaching formats such as case discussions, experiential activities, and critical writing, students can be guided to discern the essential differences between AI interaction and real social interaction, and cultivate a sober awareness of the limits of AI companionship.

More importantly, human-AI boundary education should guide students toward self-reflection. When AI always agrees with the student unconditionally, does the student lose the capacity to hear dissenting voices and confront interpersonal conflict? When AI is always available to receive the student's emotional confessions, does the student reduce active help-seeking and authentic expression toward others? This educational design—turning technological scrutiny into self-scrutiny—is crucial for cultivating students' internal boundary awareness, which constitutes the most fundamental psychological safeguard against emotional dependence.

4.2. Design principle: from "AI-first" to "reality-first"

In the design of activities for mental health education curricula, a "reality-first" principle should be established: it should prioritize guidance for students to attempt and practice real-world social support

strategies, and position AI as a supplementary option only when real-world strategies are genuinely unavailable, rather than the primary option.

This principle does not stem from technological bias but from an understanding of how social competence develops. Social skills are not declarative knowledge ("knowing what to say") but procedural competence ("being able to flexibly apply them in real situations"). The college years represent an intensive and critical period for social skill development. Every proactive attempt at real social interaction, such as introducing oneself to a new classmate, expressing a dissenting opinion in a discussion, or arranging to meet a friend, constitutes valuable skill practice. If the excessive convenience of AI continually substitutes for these practice opportunities, this developmental window may be unwittingly squandered.

At the operational level, mental health education curricula can design "graded social challenge" assignments, encouraging students to start with low-risk real-world social behaviors (e.g., having lunch with an unfamiliar classmate) and gradually progress to more challenging social tasks, rather than turning to AI for solace at the first sign of social difficulty. Concurrently, curricula should help students develop realistic expectations and tolerance for the inevitable awkwardness, misunderstandings, and clumsiness inherent in real-world social interaction, helping them understand that these "imperfections" are not signs of social failure, but rather necessary steps toward social competence.

4.3. Tool innovation: designing AI applications that promote real-world connection

Practitioners in university mental health services need not stop at boundary education and behavioral guidance. Behavioral activation offers a useful principle: AI tools should be designed to promote real-world connection, not virtual substitution.

Unlike approaches that provide virtual companionship to fill emotional voids, the core function of "connection-promoting" AI should encourage and support users in building genuine real-world interpersonal relationships. Its design should follow several principles. When the system detects that a user has been engaged in prolonged immersive interaction, it should offer a gentle boundary reminder, such as suggesting that it may be a good time to invite a friend for a walk outside, instead of continually holding the user's attention. When a user expresses loneliness, the AI should not directly provide virtual companionship as a substitute. Instead, it should ask guiding questions, such as whether there is someone nearby with whom the user could try talking. This can help the user mobilize real-world social support resources. The AI can also integrate information about campus social activities, including club recruitments, reading groups, and sports teams, and offer personalized.

Existing research proposes a three-dimensional governance framework for AI tools in mental health education, based on technology, humanities, and institutions [14]. This paper further argues that, in ethical design, AI developers should avoid anchoring users' attention to digital interfaces. Instead, real-world well-being and interpersonal connection should be the ultimate product goal. This shift marks a move from an attention economy to a well-being economy.

4.4. Systemic development: building a multi-tiered digital literacy education framework

In the longer term, universities should systematically incorporate "AI social literacy" into the content framework of general education or mandatory mental health curricula. This literacy framework comprises three key dimensions.

At the functional literacy level, students should acquire basic AI usage skills and understand the working principles, response mechanisms, and technical limitations of anthropomorphic AI. At the critical literacy level, students should develop the capacity to identify potential psychological risks in AI interaction, including

both the ability to assess the credibility of AI-generated information and awareness of the "information cocoons" and "affective cocoons" that algorithmic recommendation systems may create. At the ethical and self-care literacy level, students should be guided to establish healthy AI usage habits and boundary awareness, maintaining a balanced equilibrium between digital life and real-world life.

These three levels of literacy education should extend across the entire duration of students' undergraduate studies, rather than being confined to a single semester's course topic. The freshman orientation phase should emphasize establishing foundational cognition and boundary awareness. In the upper-year phases, education should be aligned with the social role transitions that students will soon face, further reinforcing the importance of real-world social competence and self-regulation strategies for AI use. Only when students become conscious, boundary-aware technology users can they effectively mitigate the potential erosion of social development caused by anthropomorphic AI.

5. Conclusion

The paper integrates parasocial interaction theory, self-determination theory, and uses and gratifications theory to construct a Triggering–Satisfying–Differentiating framework and systematically analyze the dual effects of anthropomorphic AI on college students' social psychology. The study finds that AI effectively alleviates loneliness and social anxiety in the short term by enhancing social presence, removing evaluative threats, and enabling immediate interaction. However, under the drive of escapist motivation, the same mechanism may translate into three latent risks: weakened motivation for real-world social interaction from substitutive gratification, elevated interpersonal expectations from idealized interaction, and reinforced avoidance from emotional dependence. For socially anxious groups, AI is particularly prone to solidify temporary avoidance into chronic withdrawal. Based on these findings, this paper proposes four response strategies: human–AI boundary education, the reality-first principle, connection-promoting AI design, and a multi-tiered digital literacy education system. These strategies are aimed at guiding students to become conscious, boundary-aware technology users, ensuring that AI serves rather than supplants the developmental value of genuine human connection.

Several limitations of this study should be acknowledged. First, the Triggering–Satisfying–Differentiating framework proposed herein is theoretical in nature and requires empirical validation. College students' AI usage motivations are often multidimensional and fluid; future research could employ latent profile analysis to identify more refined usage typologies. Second, while this study focuses on social anxiety as a vulnerability factor, other variables, such as personality traits, attachment styles, and the quality of real-world social support, may moderate the direction and intensity of AI's effects and warrant further investigation. Third, the "withdrawal" effects discussed herein are primarily supported by theoretical deduction and indirect cross-sectional evidence; future longitudinal designs could directly test causal directions and examine the implementation of the proposed educational strategies.

References

[1] Chen, J., Yuan, D., Dong, R., Cai, J., Ai, Z., & Zhou, S. (2024). AI significantly facilitates development in the mental health of college students: A bibliometric analysis. *Frontiers in Psychology, 15*, 1375294.

[2] Fu, X., Zhang, K., & Chen, X. (2023). *Report on the development of mental health of Chinese nationals (2021–2022)*. Social Sciences Academic Press.

[3] Guo, J., Wang, P., Ma, Y., Chen, L., Guo, K., Hu, Y., & Liu, H. (2026). The application of large language model-based intelligent agents in college students' psychological counseling. *Advances in Psychological*

Science, 34(6), 1072.

[4] Munnukka, J., Talvitie-Lamberg, K., & Maity, D. (2022). Anthropomorphism and social presence in human–virtual service assistant interactions: The role of dialog length and attitudes. *Computers in Human Behavior, 135*, 107343.

[5] Sun, M., Zhou, H., Li, Y., Wang, J., Yang, W., Gong, Y., Xu, J., Zhang, J., Yang, X., Bueber, M., Phillips, M. R., & Zhou, L. (2024). Professional characteristics, numbers, distribution and training of China's mental health workforce from 2000 to 2020: A scoping review. *The Lancet Regional Health–Western Pacific, 45*, 100992. https://doi.org/10.1016/j.lanwpc.2023.100992

[6] Huang, F., Ding, H., Li, S., Han, N., Di, Y., Liu, X., Zhao, N., Li, L., & Zhu, T. (2025). Self-help AI psychological counseling system based on large language models and its effectiveness evaluation. *Acta Psychologica Sinica, 57*(11), 2022–2042. https://doi.org/10.3724/SP.J.1041.2025.2022

[7] Hou, J., Zeng, H., Wang, Q., Zheng, Y., & Zhao, F. (2025). AI empowerment or hidden concern? Latent profiles of AI use motivations and the reciprocal relations between AI use motivations and mental health among university students. *Journal of Psychological Science, 48*(5), 1038–1050. https://doi.org/10.16719/j.cnki.1671-6981.20250502

[8] Gong, X., Liu, W., & Shang, Z. (2025). Ethical challenges in AI-based mental health interventions. *Academic Journal of Naval Medical University, 46*(8), 970–976.

[9] Jiao, Y., & Long, W. (2025). Application of AI in empowering mental health education in higher education institutions. *Applied & Educational Psychology, 6*(3).

[10] Thakkar, A., Gupta, A., & De Sousa, A. (2024). AI in positive mental health: A narrative review. *Frontiers in Digital Health, 6*, 1280235.

[11] Wang, Y., & Wang, H. (2024). Mediating effects of AI on the relationship between academic engagement and mental health among Chinese college students. *Frontiers in Psychology, 15*, 1477470.

[12] Ibrahim, L., Hafner, F. S., Cheng, M., Lee, C., Anselmetti, R., Willer, R., & Yang, D. (2026). Sycophantic AI makes human interaction feel more effortful and less satisfying over time. arXiv preprint arXiv:2605.07912.

[13] Gao, W., Li, Y., Yuan, J., & He, Q. (2025). The shared and distinct mechanisms underlying fear of evaluation in social anxiety: The roles of negative and positive evaluation. *Depression and Anxiety, 2025*(1), 9559056.

[14] Zhou, S., & He, Y. (2026). Model and path of generative AI empowering mental health education curriculum reform in vocational colleges. *Vocational Education, 25*(2), 76–80.