Received: 21 September 2025 | Revised: 12 December 2025 | Accepted: 21 January 2026 | Published online: 10 February 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 4(4) 520–529
DOI: 10.47852/bonviewAIA62027719



Modeling the Double-Edged Effects of AI Companions on Chinese Generation Z: A Dual-Pathway Analysis of Risk and Protection

Haiyang Li[1] and Jing Liu[1,*]

[1]*Faculty of Innovation and Design, City University of Macau, China*

Abstract: We study Xingye AI (星野·AI皆你所见), a generative emotional companion used by Chinese Gen Z. This research took advantage of a quantitative research approach by employing survey as the data collection method and descriptive statistics as the data analysis method. With survey data from N = 402 active users, we tested a dual-path model: two risk factors (bedtime procrastination and problematic use) and two protection factors (anthropomorphism and perceived empathy). We also used hierarchical regression models. Risks align with higher depression, anxiety, loneliness, and sleep disturbance; protections show the opposite pattern. Adding protections improves model fit across outcomes ($\Delta R^2 \approx 0.06$–0.09). Beyond these behavioral results, we turn them into design. The coefficients set parameters for an adaptive pacing mechanism and a lightweight bandit/RL updater that tunes empathy and timing per user. The loop keeps a supportive tone by day and discourages over-engagement after 23:00. This links measurement to implementation and points to well-being-oriented affective computing.

Keywords: AI companions, affective computing, Gen Z, adaptive pacing, digital well-being

1. Introduction

Large language model (LLM) companions are now common among young people. They are always on, remember prior chats, and can offer comfort. This makes them useful during study breaks and late-night unwinding. The same features can also lead to more checking, later bedtimes, and poorer sleep [1, 2]. They also create a strong sense of social presence that encourages reliance [3].

Two behavioral risks are central here. Bedtime procrastination is the delay of going to bed despite feeling tired [2, 4]. Problematic use refers to compulsive or hard-to-control engagement measured with short, validated tools [5]. Both can appear when a companion stays available late at night.

Two affective features may protect users. Anthropomorphism—seeing the system as more human-like—can increase trust and social presence [6]. Anthropomorphism has been widely discussed as a key factor shaping users' perceptions and interactions with intelligent systems [7]. Perceived empathy—feeling understood and supported—is linked to better coping and well-being [8]. These ideas align with broader accounts of motivation and life evaluation [9]. In conversational AI, design choices that raise empathy or human-likeness may offset risk [10, 11].

Most studies treat risks and protections separately. Few test them together in one model for mental health and sleep outcomes in companion use. Recent work on generative AI companions notes "double-edged" effects and calls for integrated tests [12]. We address this gap using Xingye AI with Chinese Gen Z users. We test a dual-path model that includes two risks (bedtime procrastination and problematic use) and two protections (anthropomorphism and perceived empathy) and relates them to depression, anxiety, loneliness, sleep disturbance, and subjective well-being [13, 14]. We use brief, validated instruments common in behavioral health and human–AI interaction (HAI) research [4, 5, 8, 15]. We examine correlations and then estimate hierarchical regressions to see whether protections add explanatory power beyond risks.

Contributions. This study makes three contributions that integrate empirical modeling with implementable design principles for AI companions.

1) An integrated, quantitative test of risks and protections within one model for LLM companions used by Gen Z. Through this study, we developed four constructs (bedtime procrastination, problematic use of AIs, anthropomorphism, and perceived empathy) into an empirical framework. These four constructs collectively account for how Gen Z user's experience mental illness, social isolation and loneliness, and well-being and sleep disturbance from daily interaction with AI companions. This framework allows a clearer understanding of both how risk factors are established and how protective factors are identified via regular interaction with AI companions.
2) A clear measurement package with brief, validated instruments that supports comparability and low burden [4, 5, 8, 15]. Alongside standard regression analysis, we translate psychological factors into input data for determining pace, controlling availability, adjusting tone, and calibrating persona as a means of assessing the interaction between users and the AI companion. By applying the standard regression coefficients, we can provide users of the AI companion with a matrix of numerical weights, which allow users to translate validated evaluations into actionable means for enhancing their well-being through the use of the AI companion.

*Corresponding author: Jing Liu, Faculty of Innovation and Design, City University of Macau, China. Email: jingliu@cityu.edu.mo

© The Author(s) 2026. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

3) Design implications that link the constructs to pacing, tone-shaping, and empathy modules in companion systems (Figures 1–3), aligning with well-being-oriented affective computing [12, 16]. The results from the empirical studies thus far have not yet been incorporated into a functioning system but rather provide a basis for developing more usage-based adaptive pacing personalization in the future. The relative magnitudes of the coefficients will indicate which of the aforementioned variables (throttling or smoothing) and/or when to adjust (nighttime or daytime) will indicate how to prioritize them. Model gain results (ΔR^2 = 0.06–0.09) further show that incorporating protective factors substantially improves prediction beyond risk-only approaches.

2. Literature Review

Conversational AI companions have moved into everyday use. They keep context, reply in natural language, and create a felt sense of "being with someone," especially in mobile and late-night settings [3, 11]. At the same time, adjacent media and sleep research warns that frequent checking and nighttime sessions relate to emotional exhaustion and poorer sleep. Recent reviews and meta-analyses confirm robust links between electronic-media use and reduced sleep quality [17]. This background frames a field with double-edged potential. Relevant studies [16, 18–22] synthesize many current academic works on AI's application to the user experience of VR, detection/recognition of scene text (OCR), management of supply chains, readiness for Industry 4.0, and predictive analytics (wine quality, HR demand, etc.) and discuss the key methods and trends within those areas, as well as research gaps that were not filled by the authors themselves.

Current evidence tends to follow two strands. One examines behavioral self-regulation and uses compact constructs to capture dysregulated engagement. Bedtime procrastination means going to bed later than intended and is tied to poorer sleep and next-day functioning across settings [4]. Problematic use refers to compulsive or hard-to-control engagement measured with brief validated tools adapted from social and smartphone contexts [5]. Both patterns become more likely when availability extends into the night, keeping conversations active and pushing bedtimes later [23]. The other strand examines relational support. Anthropomorphism can raise social presence, trust, and willingness to rely on the agent, while perceived empathy—feeling understood and supported—relates to coping and well-being [24, 25]. These findings align with broader accounts linking supportive interactions to higher well-being and life evaluation [9]. Recent work on generative companions therefore frames effects as double-edged and calls for integrated tests that place risks and protections in the same structure [12].

In practice, modern companions run an LLM with memory/persona layers that shape tone and style. They are governed by availability and pacing policies that decide when and how often replies are sent [3, 11, 24]. High availability can push conversations into the night and produce bursty exchanges, amplifying the sleep risks identified in media and sleep literature [7]. To manage this trade-off, many deployed systems adopt lightweight online personalization: dialogue-policy learning in reinforcement learning and contextual bandits adjust tone and pacing from ongoing signals via small, incremental updates [26–28]. Reviews of empathic agents add that reviews report improvements in user experience in mental health contexts while also cautioning against over-engagement and over-reliance [29].

User reports echo this mix of comfort and caution. People value warmth, continuity, and a clear sense of presence, especially when empathy cues are salient [29]. Trust is conditional: human-like style can help, but scripted or "too perfect" replies can feel inauthentic; anthropomorphism can raise expectations and, when mismatched, invite doubt [30]. Some users worry about spillover into offline life—such as displaced attention in close relationships—consistent with evidence on distraction in intimate settings and thin work–life boundaries [31]. Requested improvements are broadly consistent: clearer nighttime boundaries, adjustable availability, transparent tone/persona settings, and brief supportive responses rather than long late-night chats [23, 32]. Preferences vary by age and culture; we therefore treat warmth and boundaries as design levers rather than fixed prescriptions.

Despite these advances, many studies still analyze risks or protections in isolation, limiting inference about their relative roles for mental health and sleep in the same population and setting [5, 8]. Affective-computing research offers measurement and design frames, but fewer studies close the loop from validated constructs to implementable policies for pacing and tone in companion systems [33, 34]. Guided by these strands and user reports, the present study selects two behavioral risks—bedtime procrastination and problematic use—and two affective protections—anthropomorphism and perceived empathy—measured with brief, validated instruments common in behavioral health and HAI [4, 5, 8, 15]. This selection allows an integrated test of their relative roles for depression, anxiety, loneliness, sleep disturbance, and subjective well-being in one model while keeping respondent burden low [13]. It also prepares a bridge from measurement to implementable design: parameters that inform availability and pacing at night and calibrate tone and empathy toward well-being goals [26, 27].

Figure 1 presents the conceptual framework used in this study. Risk factors on the left (problematic use and bedtime procrastination) and protective factors on the right (anthropomorphism and perceived empathy) connect to outcomes at the center (depression, anxiety, loneliness, sleep disturbance, and subjective well-being). Potential moderators appear at the top (digital boundary, adult attachment, and peer/school support). In the present analyses, we test the risk and protection paths; moderator influences are theorized for context and future work.

This review motivates the dual-path model and the hypotheses that follow and keeps a clear line from prior evidence to measures, analyses, and later design implications.

3. How Gen Z Use Emotional-AI Companions: Sample, Measures, and Design Hooks

3.1. Research questions and hypotheses

We examine whether a dual-path model can balance help and risk in everyday companion use among Chinese Gen Z. The overarching question is how LLM companions can sustain digital well-being by combining emotional support with behavioral self-regulation. We test three research questions:

(RQ1) How the behavioral risks—bedtime procrastination and problematic use—relate to depression, anxiety, loneliness, and sleep disturbance.

(RQ2) How the affective protections—anthropomorphism and perceived empathy—relate to those outcomes plus subjective well-being.

(RQ3) Whether protections add explanatory power beyond risks.

From prior work, we derive five hypotheses:

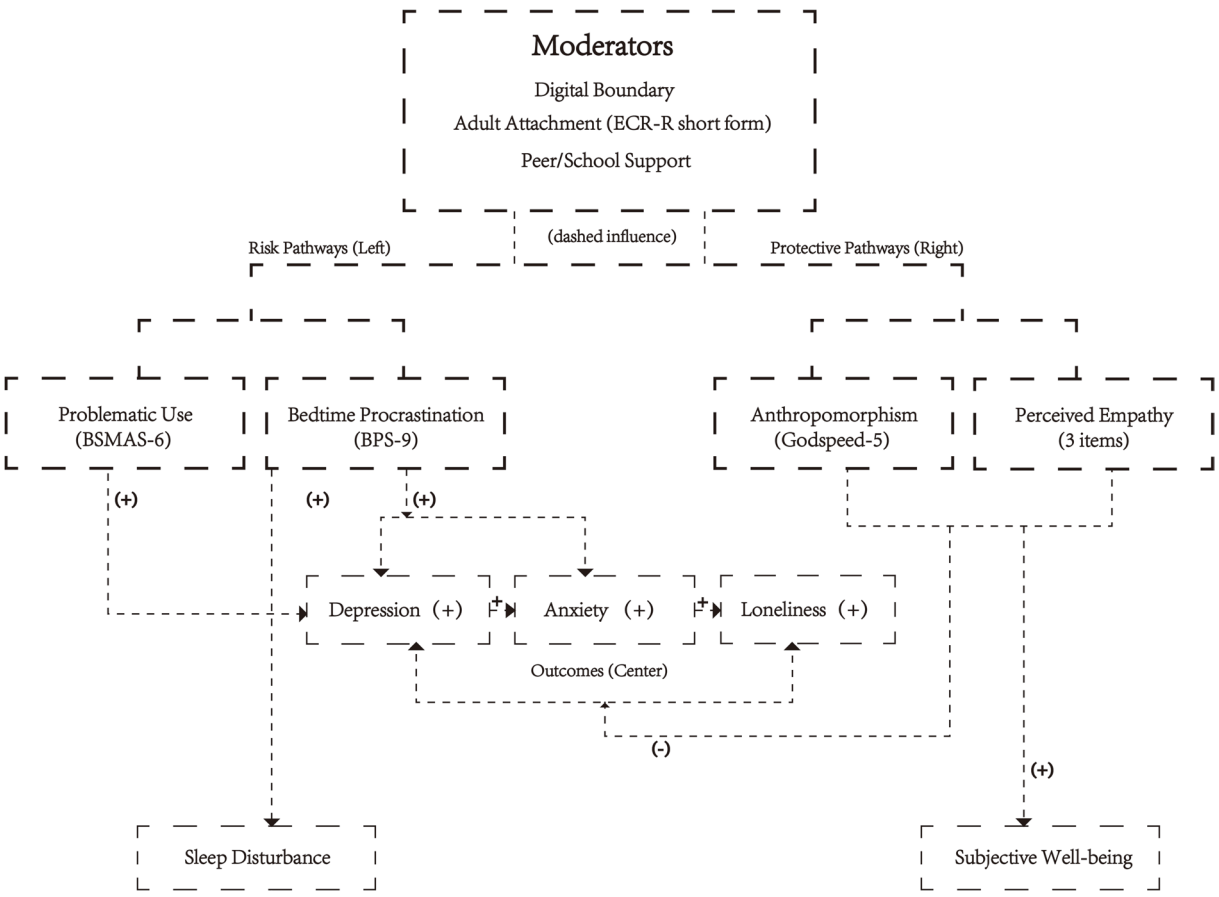
H1: higher bedtime procrastination predicts higher depression, anxiety, and sleep disturbance and lower well-being [2, 4].

H2: higher problematic use predicts higher depression, anxiety, and loneliness and poorer sleep [5, 17, 35].

H3: higher anthropomorphism predicts lower depression and loneliness, better sleep, and higher well-being [6, 24, 30].

H4: higher perceived empathy predicts lower depression and anxiety, less loneliness, better sleep, and higher well-being [8, 9, 36, 37].

Figure 1
Conceptual structure



H5: adding anthropomorphism and perceived empathy improves model fit across outcomes beyond risks [12].

3.2. Sample and procedure

We ran an online survey with N = 402 active users of Xingye AI who reported use within the past month. Participation was voluntary and anonymous; respondents gave informed consent and completed the questionnaire in one sitting. Screening removed obvious non-users; standard quality checks (attention checks, excessive speed, and straight-lining) were applied before analysis. Table 1 summarizes the demographic and usage profiles (age, gender, education/employment, daily companion time, and after 23:00 use), which reflect heavy mobile use among Chinese Gen Z.

3.3. Measures

Behavioral risks used brief, validated instruments: bedtime procrastination (BPS-9) [4] and problematic use (BSMAS-6) adapted from social/smartphone contexts [35]. Affective protections were anthropomorphism measured with the Godspeed anthropomorphism items [6] and perceived empathy with three short items drawn from empathy research [8]. Outcomes were depression (PHQ-9) [15], anxiety (GAD-7) [13], loneliness (8-item short form), sleep disturbance (three brief items consistent with sleep-risk work [2]), and subjective well-being (life-evaluation items [37]). All scales were scored as means so that higher values indicate more of the construct (higher PHQ-9/GAD-7 reflect more symptoms; higher SWB reflects greater well-

Table 1
Demographic and usage characteristics (N = 402)

Variable	Category/metric	n (%)/mean (SD)
Gender	Female	215 (53.5%)
Gender	Male	187 (46.5%)
Age (years)	Mean (SD)	21.51 (2.11)
Status	Undergraduate	175 (43.5%)
Status	Employed	146 (36.3%)
Status	Unemployed/Intern	45 (11.2%)
Status	Graduate	36 (9.0%)
Daily AI-companion use	1–2 h/1–2 h	125 (31.1%)
Daily AI-companion use	30–60 min/30–60 min	121 (30.1%)
Daily AI-companion use	15–30 min/15–30 min	71 (17.7%)
Daily AI-companion use	>2 h	62 (15.4%)
Daily AI-companion use	<15 min	23 (5.7%)
After 23:00 usage	Sometimes	135 (33.6%)
After 23:00 usage	Often	114 (28.4%)
After 23:00 usage	Seldom	79 (19.7%)
After 23:00 usage	Always	47 (11.7%)
After 23:00 usage	Never	27 (6.7%)

Table 2
Descriptive statistics and reliability of study variables

Scale	k	Observed Min–Max	Mean	SD	Cronbach's α	N
Bedtime procrastination (BPS-9)	9	1.11–5.00	3.38	1.18	0.941	402
Problematic use (BSMAS-6)	6	1.00–5.00	3.44	1.17	0.904	402
Anthropomorphism (5)	5	1.00–6.00	3.41	1.40	0.910	402
Perceived empathy (3)	3	1.00–5.00	2.61	1.23	0.833	402
Depression (PHQ-9)	9	1.00–3.00	2.18	0.66	0.930	402
Anxiety (GAD-7)	7	1.00–3.00	2.23	0.64	0.898	402
Loneliness (8 items)	8	1.12–5.00	3.44	1.21	0.936	402
Well-being (SWB-3)	3	1.00–5.00	2.52	1.19	0.820	402
Sleep disturbance (3)	3	1.00–5.00	3.42	1.25	0.824	402
Attachment (4)	4	1.25–7.00	4.65	1.40	0.874	402
Digital boundary (3)	3	1.00–5.00	2.64	1.27	0.836	402
Peer/school support (2)	2	1.00–5.00	2.62	1.35	0.801	402

being). Internal consistency coefficients are shown in Table 2 and meet accepted thresholds for research use; observed ranges indicate good variability for modeling. Given prior links between late-night engagement and sleep risks [2, 38], we also flagged after 23:00 use for descriptive checks and later design translation.

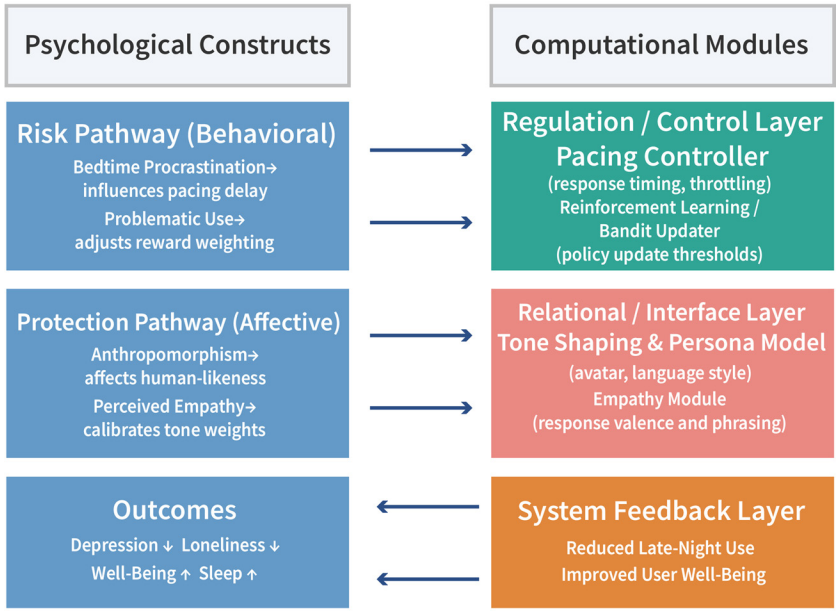
3.4. Analytic strategy

Analyses proceeded in two steps. First, we reported zero-order correlations among all focal variables to show bivariate associations. Second, for each outcome (depression, anxiety, loneliness, subjective well-being, and sleep disturbance), we estimated hierarchical regressions: Step 1 entered the two behavioral risks (BPS-9 and BSMAS-6), and Step 2 added the two protections (anthropomorphism and perceived empathy) to test incremental validity (ΔR^2). Coefficients are standardized β with 95% confidence intervals; assumptions were checked via residual diagnostics. Analyses used complete cases; Ns for each model/table are reported. Correlation tables and model summaries appear in Sections 4.1–4.2.

3.5. Robustness and design hooks

To assess robustness, we re-estimated models under alternative anthropomorphism encodings (retain full range vs. top-code at the maximum) and with light winsorizing of extreme values; qualitative patterns were expected to remain stable and are reported alongside main results. As a bridge from measurement to implementation, we map psychological constructs to computational modules—tone shaping and persona for empathy, and a pacing controller for availability/throttling—so coefficients can directly inform nighttime boundaries and online personalization in practice. Figure 2 presents this mapping and sets up Section 5 on system design.

Figure 2
Psychological constructs mapped to computational modules



4. What the Data Say: Behavioral Risks vs. Affective Protections

4.1. Descriptives and zero-order patterns

Descriptive statistics and reliability coefficients for all variables are presented in Table 2. All scales demonstrated acceptable to excellent internal consistency (α = 0.82–0.94), confirming suitability for multivariate modeling. The observed score ranges indicate substantial variability across behavioral risks, affective protections, and outcome measures, ensuring sufficient dispersion for regression analyses.

Zero-order correlations (Table 3) provide an initial test of the dual-pathway structure. Behavioral risk factors—bedtime procrastination and problematic use—showed consistent positive associations with psychological distress (PHQ-9 and GAD-7), loneliness, and sleep disturbance, and negative associations with subjective well-being (e.g., PHQ-9 r = 0.40; SWB r = −0.40). These patterns align with prior research on nighttime dysregulation and compulsive digital engagement.

In contrast, both affective protection factors—anthropomorphism and perceived empathy—were associated with lower distress and sleep disturbance and higher well-being (e.g., anthropomorphism with PHQ-9 r = −0.48; with SWB r = 0.45). Anthropomorphism exhibited the strongest protective correlations across outcomes, whereas empathy showed smaller but directionally consistent effects. Together, these preliminary associations illustrate the theoretical logic of the dual-pathway model: behavioral risks cluster with poorer mental health outcomes, while affective perceptions cluster with better well-being.

4.2. Hierarchical models

To evaluate the distinct contributions of behavioral risks and affective protections, hierarchical regression models were constructed for all five outcome variables, with the first step including bedtime procrastination and problematic use, and the second step adding anthropomorphism and perceived empathy. The initial step of each model explained a substantial proportion of variance across the outcomes, with R^2 values ranging from 0.21 to 0.30. Bedtime procrastination consistently emerged as the stronger behavioral predictor, displaying pronounced associations with anxiety (β = 0.33) and sleep disturbance (β = 0.35), while problematic use showed meaningful links to emotional indicators, particularly loneliness (β = 0.25). These patterns suggest that dysregulated engagement functions as an important antecedent of both emotional distress and disrupted sleep.

When the affective protection variables were added in the second step, the models exhibited clear improvements in explanatory power, with increases in explained variance between 0.06 and 0.09 across all outcomes (see Table 4), with coefficients ranging from −0.22 to −0.28 for depression and anxiety, +0.24 for subjective well-being, and −0.20 for sleep disturbance. These results indicate that perceiving the AI companion as more human-like is consistently associated with lower distress and better psychological functioning. Perceived empathy also contributed additional protective effects, with coefficients generally falling between 0.12 and 0.17 in absolute value, demonstrating that the feeling understood by the AI companion offers incremental benefits beyond those associated with anthropomorphism.

The addition of anthropomorphism and empathy attenuated the coefficients for bedtime procrastination and problematic use while preserving their significance. This attenuation suggests that the affective pathway accounts for part of the variance previously attributed to behavioral risks, yet the persistence of significant effects indicates that both pathways independently exert meaningful influences on user outcomes. Overall, the hierarchical results support the theoretical expectation that behavioral dysregulation and affective perceptions jointly shape psychological and sleep-related responses, and that the simultaneous inclusion of both pathways yields a more complete and accurate representation of the data than either pathway considered alone.

Table 3
Zero-order correlations among predictors and outcomes

Predictor/ outcome	PHQ-9	GAD-7	UCLA8	SWB-3	Sleep3
BPS-9	0.40	0.45	0.49	-0.40	0.45
BSMAS-6	0.36	0.44	0.41	-0.37	0.40
Anthro-5	-0.48	-0.47	-0.50	0.45	-0.46
Empathy-3	-0.39	-0.40	-0.45	0.38	-0.43

4.3. Robustness

To better quantify each predictor's contribution, we compared standardized coefficients across outcomes. Anthropomorphism emerged as the strongest single predictor in the model—stronger than both risk variables for depression, loneliness, and well-being—highlighting its central role in socio-emotional buffering. Bedtime procrastination was the strongest risk predictor overall, particularly for sleep-related outcomes.

Table 4
Hierarchical regressions predicting outcomes

Outcome	Model	BPS-9 (β)	BSMAS-6 (β)	Anthro (β)	Empathy (β)	R^2	ΔR^2
PHQ-9	1	0.30	0.24			0.21	
PHQ-9	2	0.16	0.12	-0.28	-0.14	0.29	0.09
GAD-7	1	0.33	0.30			0.28	
GAD-7	2	0.22	0.20	-0.22	-0.12	0.34	0.06
UCLA8	1	0.39	0.25			0.30	
UCLA8	2	0.26	0.13	-0.24	-0.17	0.38	0.08
SWB-3	1	-0.30	-0.25			0.21	
SWB-3	2	-0.18	-0.14	0.24	0.13	0.28	0.07
Sleep3	1	0.35	0.26			0.26	
Sleep3	2	0.23	0.15	-0.20	-0.17	0.33	0.07

Model-gain indices (ΔR^2) revealed that including affective protections improved explained variance by 6–9 percentage points, which corresponds to a 29%–43% relative increase over risk-only models. Gains were largest for depression and loneliness, indicating that socio-emotional mechanisms (e.g., perceived warmth and human-likeness) are especially influential for emotional outcomes. These analyses deepen the empirical support for the dual-pathway model and demonstrate that affective protections are not merely supplementary—they substantially enhance prediction accuracy.

4.4. Outcome-by-outcome narrative

Across outcomes, patterns are consistent. For depression and anxiety, both risks are positively related; adding protections reduces coefficients for risks and introduces sizable negative effects for anthropomorphism and smaller, aligned effects for empathy. Loneliness shows the same structure: risks predict more loneliness, and protections predict less, with anthropomorphism the stronger signal. For subjective well-being, signs reverse as expected—risks relate to lower well-being while protections relate to higher well-being—and the addition of protections yields one of the larger improvements in fit. Sleep disturbance follows the behavioral logic: bedtime procrastination and problematic use relate to worse sleep; protections relate to better sleep (smaller disturbance) even after controlling for risks.

These results answer the research questions directly. RQ1 is supported: bedtime procrastination and problematic use are reliably associated with higher distress and poorer sleep, and lower well-being (Table 3; Step-1 models), consistent with H1–H2. RQ2 is supported: anthropomorphism and perceived empathy predict better mental-health, higher well-being, and less sleep disturbance, with anthropomorphism typically the larger protection; effects remain after accounting for risks (Table 5), supporting H3–H4. RQ3 is supported: adding protections improves model fit for every outcome ($\Delta R^2 \approx 0.06$–0.09), demonstrating incremental validity beyond risks (H5).

Taken together, both paths matter: risks track dysregulated engagement (late-night delays and compulsive bursts), while protections capture supportive, human-like interaction. Anchored in Figure 1 and operationalized via the mapping in Figure 2, the coefficients translate into implementable controls—nighttime boundaries and online personalization. Figure 3 sketches the adaptive loop that keeps a warm tone by day while gently throttling engagement after 23:00, setting up Section 5.

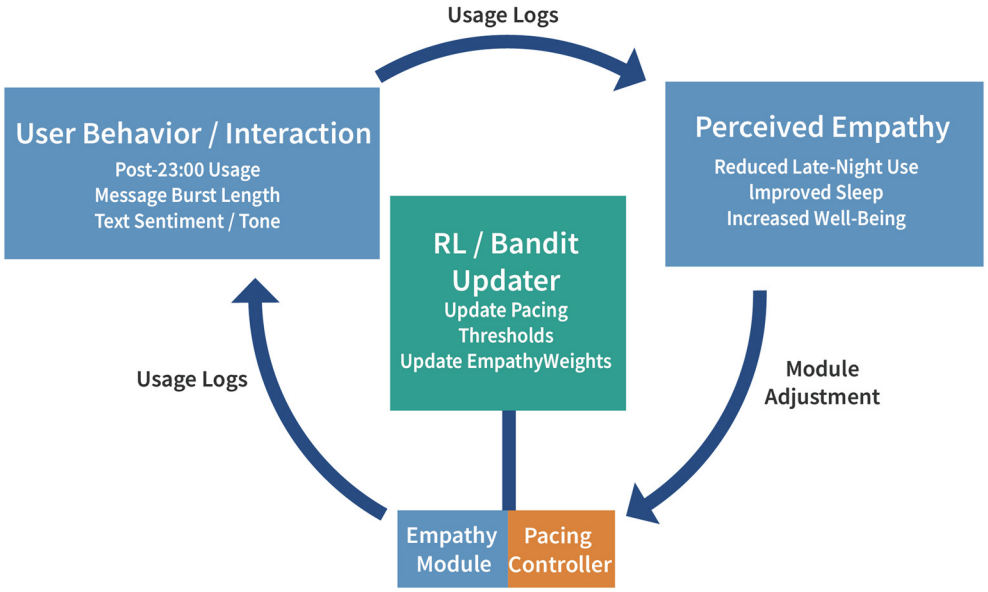
4.5. Integrated interpretation of outcomes

Across all outcomes, a clear asymmetry emerged between the behavioral and affective pathways. Behavioral risks were more strongly associated with physiological and arousal-related variables, particularly sleep disturbance and anxiety, whereas affective protection factors were more influential in predicting socio-emotional outcomes such as depression, loneliness, and subjective well-being. This pattern aligns with theoretical accounts in which dysregulated nighttime engagement disrupts emotional and physiological stability, whereas perceptions of warmth, understanding, and human-like presence buffer

Table 5
Robustness of anthropomorphism correlations to alternative "6" encodings

Outcome	r (keep 6)	r (6→missing)	r (6→5)	N (keep)	N (miss)	N (5)
PHQ-9	-0.475	-0.478	-0.485	402	399	402
GAD-7	-0.470	-0.497	-0.490	402	399	402
UCLA8	-0.503	-0.501	-0.513	402	399	402
SWB-3	0.451	0.472	0.467	402	399	402
Sleep3	-0.458	-0.452	-0.465	402	399	402

Figure 3
Adaptive pacing loop with bandit/RL updater



negative experiences and promote well-being. Overall, the results provide coherent empirical support for the dual-pathway model and demonstrate that considering both risk and protection mechanisms yields a substantially more accurate depiction of user outcomes than examining either pathway alone.

5. Making It Work: Nighttime Boundaries and Online Personalization

5.1. Adaptive control loop and nighttime boundaries

Building on the adaptive loop introduced above, the system tunes two per-user levers—pacing (reply timing/throttling) and empathy weights (tone/length)—from ongoing usage signals while respecting local time and safety constraints. Inputs include after-23:00 activity share, burst length, check frequency, and recent sentiment/tone; state features add standardized BPS-9, BSMAS-6, Anthro-5, and Empathy-3 scores plus time-of-day. At each step, the updater selects an action (pacing tier and empathy tier) and optimizes a reward proxy: reduce after-23:00 activity and bursts without degrading next-day sentiment or session quality. We use a conservative contextual-bandit update (ε-greedy or UCB) with per-user value estimates and weekly decay; exploration is capped at night. Safety rails apply: a hard cap on delay, a minimum empathy floor, instant rollback if complaint/help-seeking signals rise, and crisis cues that bypass pacing entirely.

1) Nighttime boundaries (gentle, not blocking).
2) Tiered delays. Map a user's risk index to 5–15–30 s delay bands; auto-decay to lower tiers as risk subsides.
3) Burst smoothing. When rapid-fire turns occur, insert micro-pauses and hand off with a "pick up tomorrow" summary.
4) Soft handovers. Optional sleep mode ("I'll check in at 08:30"), always with a one-tap opt-out.
5) Tone shift, not content loss. Keep warmth but use shorter, calmer, plan-oriented replies at night; avoid energizing prompts.
6) Local time and exceptions. Honor travel and declared night-shift schedules; expose user controls to adjust quiet hours.
7) Transparency. A one-line banner explains late-night pacing with a link to settings and details.
8) These boundaries implement Section 4's pattern: risks concentrate in late-night dysregulation, while gentle pacing curbs over-engagement without removing support.

5.2. Online personalization: pacing and tone, learned cautiously

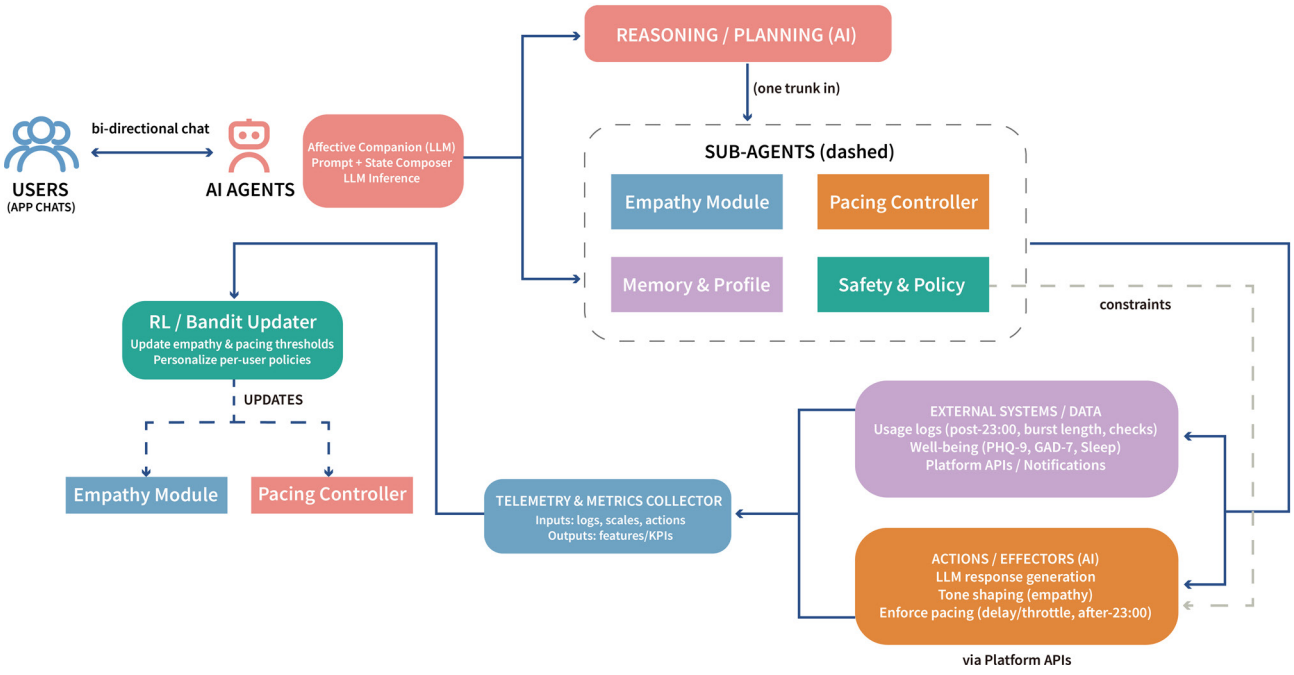
The online updater reads a compact set of signals—after-23:00 activity share, burst length, check frequency, recent sentiment and tone, and the standardized scores for BPS-9, BSMAS-6, Anthro-5, and Empathy-3—together with time-of-day. From these, it chooses one of a few pacing levels and one of a few tone/empathy levels. The objective is straightforward: reduce late-night activity and bursts without hurting next-day sentiment or session quality. We use a conservative contextual-bandit policy (ε-greedy or UCB) with per-user value estimates and weekly decay, and we cap exploration at night. Safety rails apply throughout: delays have a hard maximum, empathy never drops below a supportive floor, any rise in complaints or help-seeking triggers an immediate rollback, and crisis language bypasses pacing entirely. In practice, this means that the system stays warm and substantive during the day—brief affirmation, reflective rephrasing, and concrete next steps—and shifts at night to short, calm, plan-oriented replies that avoid energizing prompts while keeping support available.

5.3. Deploying safely at scale

To move from a study to a product, the loop sits inside a transparent, auditable stack (see Figure 4).

We track four indicators: the share of after-23:00 messages and burst length should fall, next-day sentiment and engagement should stay at or above baseline, and opt-out and complaint rates should not rise. Rollout uses stratified randomization, night-only treatment, staggered waves, and cluster guardrails to prevent unintended global shifts. Fairness and explainability are built in: we audit effects by

Figure 4
System-level architecture for emotional-AI companions



locale, gender, and schedule (including night-shift users), we provide an always-available pause/override, and we show a one-line explanation when pacing is active. Privacy and transparency follow a minimum-data stance—prefer on-device time detection, keep logs lean, and expose clear settings for pacing and data use. If messages suggest acute risk, pacing is bypassed and the established crisis protocol is triggered.

6. Discussion and Design Implications

The findings demonstrate that behavioral and affective mechanisms jointly shape how Gen Z users experience AI companions. Bedtime procrastination and problematic use predicted distress and sleep disturbance, whereas anthropomorphism and perceived empathy served as protective factors that improved well-being. The protective path added explanatory power beyond behavioral risks, confirming the "double-edged" nature of emotional-AI use while showing that empathic and human-like features can buffer against the dysregulation caused by constant availability. The results from this research reinforce and lend quantifiable evidence to previous theories about how much risk and protective factors affect the ability to self-regulate, aided by emotional support, through interaction with AI.

This research adds to psychological theories about self-regulation and empathy by examining how self-regulation and empathy manifest in AI environments. Procrastinating on going to bed can be considered an inability to manage your time during periods when digital devices become an intrusion on your life; however, the use of empathy and anthropomorphism may help users meet their needs for relatedness and competence in restoring balance. In addition, the need for authenticity: if AI provides responses that are excessively scripted or "perfect," they will diminish the user's trust. Conversely, if the designers demonstrate calibrated empathy toward users, the users will experience feeling connected to the designer without intruding on their personal boundaries. Hence, the well-being of users who receive companionship from AI will be better influenced by how warm and respectful the designers have demonstrated those attributes to their users, rather than the amount of time the users have interacted with the designers.

From a computational standpoint, where coefficients are estimates, they provide coefficients for designers to use in adapting their AI system design to find the correct balance of pacing and empathy, within the context of the user's current situation and feedback control system using low-complexity contextual bandit or reinforcement learning. Therefore, depending on the state of the user at that moment, designers should use gentle pacing for AI conversations to initiate after 23:00, promote empathy through short and calm responses, and explain the rationale for pacing AI conversations, to increase user confidence. Although the data are cross-sectional and limited to one demographic, the framework illustrates how validated psychological constructs can inform deployable affective-computing modules and how emotional-AI companions can promote healthier digital routines aligned with user well-being [39]. Recent research has emphasized the importance of designing AI systems that explicitly support user well-being and healthy digital routines [39].

This research offers three primary contributions within this broader discussion. First, it introduces an integrated dual-pathway model that unifies behavioral risks and affective protections in AI-companion use. Second, it shows how psychological constructs such as self-regulation, anthropomorphism, and perceived empathy can function as operational levers for computational mechanisms, informing the design of adaptive pacing, tone calibration, and boundary-setting systems in AI companions. Third, the empirical results provide a parameterized foundation for future well-being-oriented AI architectures by demonstrating that affective protections significantly enhance model performance relative to risk factors. Collectively, these contributions provide both a theoretical advancement in understanding emotional-AI interactions and a practical roadmap for designing next-generation AI companions that promote healthier, more balanced digital engagement.

7. Conclusion

Building on the discussion above, this section summarizes the key contributions and limitations of the study and outlines directions for future research. We tested a dual-path model for LLM companions in everyday use among Chinese Gen Z, combining two behavioral risks (bedtime procrastination and problematic use) with two affective protections (anthropomorphism and perceived empathy). Risks related to higher distress and poorer sleep, protections showed the opposite pattern, and protections added explanatory power beyond risks ($\Delta R^2 \approx 0.06$–0.09). These findings answer the three research questions and support H1–H5 in a single, integrated structure.

Our contribution is threefold. Empirically, we provide a compact, validated measurement package that compares risks and protections in the same population and outcomes. Conceptually, we sharpen the "double-edged" account by quantifying the relative roles of behavioral dysregulation and supportive, human-like interaction. Computationally, we translate coefficients into implementable controls: gentle nighttime pacing and tone calibration that keep daytime warmth while discouraging over-engagement after 23:00 through cautious online personalization.

Several limitations remain. The data are cross-sectional and self-report, drawn from one product context; causal claims cannot be made. Very short empathy items may understate protective effects. Potential moderators (digital boundary, attachment, and peer/school support) were theorized but not tested. Common-method variance and cultural specificity may limit generalizability.

Future work should use preregistered longitudinal and field experiments to test pacing versus status quo, combine telemetry with objective sleep measures, and adapt timing thresholds per user rather than a fixed 23:00 rule. Richer empathy measures and qualitative studies of authenticity can refine the protection path. Fairness audits across locale, gender, and night-shift users, together with clear user controls and transparency, will be essential for safe deployment. Taken together, this study offers a practical bridge from psychological measurement to well-being-oriented affective computing, showing how companions can remain helpful and self-regulating in everyday life.

Ethical Statement

All subjects provided informed consent for inclusion before participating in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Faculty of Innovation and Design of the City University of Macau [Reference No.: 20261201504].

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available on Zenodo at https://doi.org/10.5281/zenodo.18279792.

Author Contribution Statement

Haiyang Li: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Jing Liu:** Conceptualization, Writing – review & editing, Supervision.

References

[1] Carlson, S. E., Suchy, Y., Baron, K. G., Johnson, K. T., & Williams, P. G. (2023). A daily examination of executive functioning and chronotype in bedtime procrastination. *Sleep*, *46*(8), zsad145. https://doi.org/10.1093/sleep/zsad145

[2] Zhang, Y., Rehman, S., Addas, A., Ahmad, M., & Khan, A. (2025). The mediating role of cognitive reappraisal on bedtime procrastination and sleep quality in higher educational context: A three-wave longitudinal study. *Nature and Science of Sleep*, *17*, 129–142. https://doi.org/10.2147/NSS.S497183

[3] Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

[4] Noor, F., Rizvi, A. Z., Naveed, S., Ashraf, H., & Adeeb, M. (2025). Smartphone addiction and positive mental health of university students: Bedtime procrastination as a mediator. *Journal for Current Sign*, *3*(3), 392–407.

[5] Fernandes, H. R., Pereira, H. P., Gonçalves, M. P., Ramião, E. Z., Antunes, H. I., Fernandes, F. C., & Barbosa, I. C. (2023). Psychometric properties of the Bergen Facebook addiction scale for Portuguese adults. *The Psychologist: Practice & Research Journal*, *6*(1), 1–8.

[6] Kühne, R., & Peter, J. (2023). Anthropomorphism in human–robot interactions: A multidimensional conceptualization. *Communication Theory*, *33*(1), 42–52. https://doi.org/10.1093/ct/qtac020

[7] Makady, H. (2024). Human, I know how you feel: Individual psychological determinants influencing smartwatch anthropomorphism. *Journal of Technology in Behavioral Science*, *9*(2), 369–386. https://doi.org/10.1007/s41347-023-00351-0

[8] Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://psycnet.apa.org/doi/10.1037/0022-3514.44.1.113

[9] Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

[10] Chi, N. T. K., & Hoang Vu, N. (2023). Investigating the customer trust in artificial intelligence: The role of anthropomorphism, empathy response, and interaction. *CAAI Transactions on Intelligence Technology*, *8*(1), 260–273. https://doi.org/10.1049/cit2.12133

[11] Dailah, H. G., Koriri, M., Sabei, A., Kriry, T., & Zakri, M. (2024). Artificial intelligence in nursing: Technological benefits to nurse's mental health and patient care quality. *Healthcare*, *12*(24), 2555. https://doi.org/10.3390/healthcare12242555

[12] Smith, M. G., Bradbury, T. N., & Karney, B. R. (2025). Can generative AI chatbots emulate human connection? A relationship science perspective. *Perspectives on Psychological Science*, *20*(6), 1081–1099. https://doi.org/10.1177/17456916251351306

[13] Han, Z. R., Fang, H., Ahemaitijiang, N., & Wang, H. (2025). Generalized anxiety disorder scale (GAD-7). In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of assessment in mindfulness research* (pp. 1745–1760). Springer. https://doi.org/10.1007/978-3-031-47219-0_87

[14] Ahmed, O., Walsh, E. I., Dawel, A., Alateeq, K., Oyarce, D. A. E., & Cherbuin, N. (2024). Social media use, mental health and sleep: A systematic review with meta-analyses. *Journal of Affective Disorders*, *367*, 701–712. https://doi.org/10.1016/j.jad.2024.08.193

[15] Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

[16] Shang, W., & Alena, K. (2025). Advancements in virtual reality technology: A systematic review of user experience and application trends. *Artificial Intelligence and Applications*, *3*(4), 341–358. https://doi.org/10.47852/bonviewAIA52024327

[17] Zhuang, J., Mou, Q., Zheng, T., Gao, F., Zhong, Y., Lu, Q., …, & Zhao, M. (2023). A serial mediation model of social media addiction and college students' academic engagement: The role of sleep quality and fatigue. *BMC Psychiatry*, *23*(1), 333. https://doi.org/10.1186/s12888-023-04799-5

[18] Pal, U., Halder, A., Shivakumara, P., & Blumenstein, M. (2024). A comprehensive review on text detection and recognition in scene images. *Artificial Intelligence and Applications*, *2*(4), 229–249. https://doi.org/10.47852/bonviewAIA42022755

[19] Goswami, S. S., Mondal, S., Sarkar, S., Gupta, K. K., Sahoo, S. K., & Halder, R. (2025). Artificial intelligence-enabled supply chain management: Unlocking new opportunities and challenges. *Artificial Intelligence and Applications*, *3*(1), 110–121. https://doi.org/10.47852/bonviewAIA42021814

[20] Saleh, N. I., & Ijab, M. T. (2025). A systematic literature survey on the role of artificial intelligence techniques in Industrial Revolution 4.0 readiness. *Artificial Intelligence and Application*, *3*(3), 236–251. https://doi.org/10.47852/bonviewAIA2202336

[21] Niyogisubizo, J., de Dieu Ninteretse, J., Nziyumva, E., Nshimiyimana, M., Murwanashyaka, E., & Habiyakare, E. (2025). Towards predicting the quality of red wine using novel machine learning methods for classification, data visualization, and analysis. *Artificial Intelligence and Applications*, *3*(1), 31–42. https://doi.org/10.47852/bonviewAIA42021999

[22] Gupta, S. K. (2025). An effective opinion mining-based K-nearest neighbors algorithm for predicting human resource demand in business. *Artificial Intelligence and Applications*, *3*(3), 252–261. https://doi.org/10.47852/bonviewAIA42022379

[23] Han, X., Zhou, E., & Liu, D. (2024). Electronic media use and sleep quality: Updated systematic review and meta-analysis. *Journal of Medical Internet Research*, *26*, e48356. https://doi.org/10.2196/48356

[24] Maeda, T., & Quan-Haase, A. (2024). When human-AI interactions become parasocial: Agency and anthropomorphism in affective design. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1068–1077. https://doi.org/10.1145/3630106.3658956

[25] Wang, X., Xie, X., Wang, Y., Wang, P., & Lei, L. (2017). Partner phubbing and depression among married Chinese adults: The roles of relationship satisfaction and relationship length. *Personality and Individual Differences*, *110*, 12–17. https://doi.org/10.1016/j.paid.2017.01.014

[26] Kwan, W.-C., Wang, H.-R., Wang, H.-M., & Wong, K.-F. (2023). A survey on recent advances and challenges in reinforcement learning methods for task-oriented dialogue policy learning. *Machine Intelligence Research*, *20*(3), 318–334. https://doi.org/10.1007/s11633-022-1347-y

[27] Ban, Y., Qi, Y., & He, J. (2024). Neural contextual bandits for personalized recommendation. In *Companion Proceedings of the ACM Web Conference 2024*, 1246–1249. https://doi.org/10.1145/3589335.3641241

[28] Gan, M., & Kwon, O.-C. (2022). A knowledge-enhanced contextual bandit approach for personalized recommendation

in dynamic domains. *Knowledge-Based Systems*, *251*, 109158. https://doi.org/10.1016/j.knosys.2022.109158

[29] Shen, J., DiPaola, D., Ali, S., Sap, M., Park, H. W., & Breazeal, C. (2024). Empathy toward artificial intelligence versus human experiences and the role of transparency in mental health and social support chatbot design: Comparative study. *JMIR Mental Health*, *11*, e62679. https://doi.org/10.2196/62679

[30] Yu, Y., Yang, Z., Sun, Z., Zhao, Z., & Fu, M. (2025). A meta-analysis of anthropomorphism of artificial intelligence in tourism. *Asia Pacific Journal of Tourism Research*, *30*(9), 1207–1225. https://doi.org/10.1080/10941665.2025.2486014

[31] Gomez-Outes, A., Alcubilla, P., Calvo-Rojas, G., Terleira-Fernandez, A. I., Suárez-Gea, M. L., Lecumberri, R., & Vargas-Castrillon, E. (2021). Meta-analysis of reversal agents for severe bleeding associated with direct oral anticoagulants. *Journal of the American College of Cardiology*, *77*(24), 2987–3001. https://www.jacc.org/doi/10.1016/j.jacc.2021.04.061

[32] Sanjeewa, R., Iyer, R., Apputhurai, P., Wickramasinghe, N., & Meyer, D. (2024). Empathic conversational agent platform designs and their evaluation in the context of mental health: Systematic review. *JMIR Mental Health*, *11*, e58974. https://doi.org/10.2196/58974

[33] Concannon, S., & Tomalin, M. (2024). Measuring perceived empathy in dialogue systems. *AI & Society*, *39*(5), 2233–2247. https://doi.org/10.1007/s00146-023-01715-z

[34] Afzal, S., Khan, H. A., Piran, M. J., & Lee, J. W. (2024). A comprehensive survey on affective computing: Challenges, trends, applications, and future directions. *IEEE Access*, *12*, 96150–96168. https://doi.org/10.1109/ACCESS.2024.3422480

[35] Holte, A. J., Aukerman, K., Padgett, R., & Kenna, M. (2024). "Let me check my phone just one more time": Understanding the relationship of obsessive-compulsive disorder severity and problematic smartphone use. *Current Psychology*, *43*(13), 11593–11603. https://doi.org/10.1007/s12144-023-05298-2

[36] Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*(2), 310–357. https://psycnet.apa.org/doi/10.1037/0033-2909.98.2.310

[37] Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

[38] Hill, V. M., Rebar, A. L., Ferguson, S. A., Shriane, A. E., & Vincent, G. E. (2022). Go to bed! A systematic review and meta-analysis of bedtime procrastination correlates and sleep outcomes. *Sleep Medicine Reviews*, *66*, 101697. https://doi.org/10.1016/j.smrv.2022.101697

[39] Pei, G., Li, H., Lu, Y., Wang, Y., Hua, S., & Li, T. (2024). Affective computing: Recent advances, challenges, and future trends. *Intelligent Computing*, *3*, 0076. https://doi.org/10.34133/icomputing.0076

How to Cite: Li, H., & Liu, J. (2026). Modeling the Double-Edged Effects of AI Companions on Chinese Generation Z: A Dual-Pathway Analysis of Risk and Protection. *Artificial Intelligence and Applications*, *4*(4), 520–529. https://doi.org/10.47852/bonviewAIA62027719