Generative AI And Self-Disclosure In Underprivileged Children: Evidence From Hiwel Learning Stations

Ritu Dangwal

Position & affiliation: Area Director, Humanities and Social Sciences, NIIT University, Neemrana, Rajasthan, India. Email id: Ritu.dangwal@niituniversity.in Orchid id: 0000-0002-6950-1108

Abstract

Conversational AI platforms such as ChatGPT and Gemini have become part of everyday life, with human-like interactions encouraging users to disclose personal thoughts and feelings with reduced fear of judgment. This study examined whether access to such platforms enhances self-disclosure among children and early adolescents (aged 6–14) from underprivileged backgrounds, using HiWEL Learning Stations, public, internet-enabled kiosks for self-directed learning.

A quasi-experimental design was conducted across two underprivileged communities in New Delhi, India. The experimental site provided access to ChatGPT and Gemini, while the control site retained the standard HiWEL interface without generative AI. Eighty children (40 per group) completed a self-disclosure scale at baseline and after three months. The experimental group scored significantly higher than the control group (mean difference = 63.80; $t(78) = 13.873$, $p < .001$), with self-disclosure rising 101.4% across all five dimensions: Relationships, Personal Matters, Beliefs, Interests, and Intimate Feelings. A subsequent human–computer trust assessment revealed moderate-to-high trust ($M = 79.41$), with competency rated highest.

These findings show generative AI can significantly enhance self-disclosure among underprivileged children in minimally supervised community learning environments, highlighting its potential for inclusive learning while underscoring the need for robust child safeguarding and AI governance.

Keywords: Self-disclosure, AI conversational agents, Anthropomorphism, HiWEL Learning Stations, Child–AI interaction

Introduction

Background

The advent of the Internet fundamentally transformed the way people communicate, access information, and build social relationships. Since the mid-1990s, communication has consistently been identified as one of the primary reasons for Internet use, enabling individuals to connect across geographical, cultural, and social boundaries. Email, discussion forums, instant messaging, blogs, and later social networking platforms gradually shifted communication from being constrained by time and location to one that was continuous, immediate, and increasingly interactive. This digital transformation reshaped interpersonal communication by allowing individuals to maintain existing relationships while simultaneously creating opportunities to establish entirely new forms of online communities and social networks.

The rapid growth of online communication also generated considerable academic debate regarding its psychological and social consequences. Early researchers questioned whether increased Internet use would strengthen or weaken social relationships and community engagement. It was argued that extensive online communication might reduce face-to-face interactions with neighbours, friends, and family members, thereby weakening social support networks, interpersonal trust, and psychological well-being. Researchers were particularly concerned that virtual communication would replace meaningful interpersonal relationships with superficial online interactions, resulting in loneliness, social isolation, and reduced participation in community life.

Many researchers worried that the ease of Internet communication might encourage people to spend more time alone, talking online with strangers or forming superficial "drive-by" relationships, at the expense of deeper face-to-face discussions and companionship with friends and family. Even when individuals used the Internet to communicate with existing social ties, scholars questioned whether online interactions would displace richer forms of face-to-face and telephone conversations (Kraut et al., 2002). These concerns were reflected in the influential Internet Paradox study (Kraut et al., 1998), which initially reported increased loneliness and declining family communication among Internet users. Interestingly, subsequent longitudinal analyses demonstrated that these negative outcomes largely disappeared as individuals became more familiar with digital technologies. Rather than replacing offline relationships, the Internet increasingly served to strengthen existing social connections while providing new opportunities for communication, collaboration, and learning (Kraut et al., 2002). This evolution illustrates that technological innovations often generate initial societal concerns, many of which change as users adapt their behaviours and integrate new technologies into everyday life.

The emergence of conversational AI has become one of the fastest digital technologies in history, with hundreds of millions of users interacting daily with LLMs like ChatGPT, Gemini, Claude, and Microsoft Copilot. Unlike previous technological innovations that supported information access, Generative AI introduces a fundamentally different mode of human-computer interaction. Instead of merely retrieving information, these systems generate responses, can hold conversations, have the ability to remember the context, adapt themselves to user preferences . Not just that but they are able to simulate empathy, curiosity, encouragement, and emotional understanding. As a result they seem to occupy a unique position between information technology and social interaction.

Platforms such as OpenAI's ChatGPT and Google's Gemini use their conversational style, first-person language, and ability to simulate human-like interactions serve to personify and anthropomorphize these systems (Strauss & Voigt, 2024; Liao & Wilson, 2024). Unlike conventional search engines that require users to formulate precise search queries and independently interpret retrieved information, conversational AI engages users in a dynamic exchange that resembles dialogue with another person. These systems ask follow-up questions, provide encouragement, acknowledge uncertainty, apologise for mistakes, and tailor responses according to the ongoing conversation. Such characteristics create an impression of social presence that extends far beyond traditional human-computer interaction.

This rapid integration of conversational AI into everyday life is raising questions remarkably similar to those posed during the early Internet era, but at a considerably deeper psychological level. Earlier debates centred on whether people would replace face-to-face communication with online communication. Contemporary concerns extend beyond communication itself to whether individuals may begin forming meaningful emotional relationships with artificial conversational partners. Increasingly, conversational AI is being used not only to obtain information but also to seek advice, emotional reassurance, companionship, creative collaboration, and psychological support. Emerging evidence suggests that many users voluntarily disclose highly personal information—including emotions, fears, aspirations, relationship concerns, and mental health issues to AI systems that they perceive as patient, non-judgmental, and constantly available.

The very nature of AI conversational agents embodies an implied subjectivity by wishing users well, expressing enthusiasm or regret, apologising for errors, and providing encouragement throughout conversations. For example, ChatGPT greets users with "How can I help you today?" while Microsoft's Copilot encourages users to "Ask me anything." Throughout extended conversations, AI systems routinely employ socially supportive expressions such as "I'm glad you shared that," "That sounds difficult," or "Feel free to tell me more." Although these responses are computationally generated rather than emotionally experienced, they nevertheless contribute to perceptions of warmth, attentiveness, and empathy. Such conversational cues encourage users to engage more openly and to share personal thoughts, emotions, preferences, and experiences, a process commonly referred to as self-disclosure. Communicating with conversational AI therefore differs fundamentally from searching the web because it provides many of the psychological characteristics associated with interpersonal communication while eliminating fears of criticism, embarrassment, rejection, or social evaluation.

These developments are particularly significant for children and adolescents. Developmental psychology recognises middle childhood and adolescence as periods characterised by identity formation, emotional development, curiosity, and increasing independence from parents. Young people actively seek opportunities to express themselves, explore personal beliefs, and discuss sensitive issues that they may hesitate to raise with adults. Conversational AI may therefore represent a uniquely attractive communication partner because it is perceived as endlessly patient, immediately responsive, and non-judgmental. However, despite rapidly increasing adoption of AI among younger users, relatively little empirical evidence exists regarding how children interact socially and emotionally with conversational AI, particularly in educational settings and among socioeconomically disadvantaged populations.

HiWEL Learning Stations, part of India's Hole-in-the-Wall project (Mitra, 2003; Mitra, 2010), provide unsupervised computer access to underprivileged children, fostering self-directed digital learning. Existing research has demonstrated their contribution to digital literacy, collaborative learning, self-regulated learning, problem-solving skills, academic achievement, and children's confidence through peer interaction and minimal adult intervention. The pedagogical philosophy underpinning HiWEL encourages children to explore technology independently, making these Learning Stations a unique naturalistic environment for examining children's interaction with conversational AI. Unlike classrooms, where teacher supervision may influence children's responses, HiWEL provides a setting in which children engage with technology voluntarily, collaboratively, and with minimal external regulation.

The present study therefore investigates whether children disclose their feelings, thoughts, personal experiences, and aspirations to ChatGPT and Gemini while using HiWEL Learning Stations. More specifically, it examines whether access to conversational AI significantly increases self-disclosure compared with traditional Internet access and explores the role of human-computer trust in this process. By focusing on underprivileged children, a population largely overlooked in current AI research this study contributes to the emerging literature on child-AI interaction while providing evidence relevant to educational practice, child safeguarding, and responsible AI governance.

Literature Review

A substantial body of research spanning over five decades has examined artificial intelligence (AI) as a conversational partner capable of engaging humans in increasingly natural interactions. The evolution of conversational AI can be traced from simple rule-based systems such as ELIZA to today's sophisticated large language models (LLMs), including ChatGPT and Gemini. ELIZA, developed by Weizenbaum (1966), was designed to simulate the conversational style of a Rogerian psychotherapist by reformulating users' statements into reflective questions. Although ELIZA relied on simple pattern-matching algorithms rather than genuine understanding, many users attributed intelligence, empathy, and intentionality to the system and engaged in surprisingly intimate conversations. This phenomenon, later referred to as the "ELIZA effect," demonstrated that even primitive computer programs could evoke emotional responses and encourage personal disclosure. Subsequent conversational systems such as PARRY (Colby et al., 1972), which simulated the conversational patterns of an individual with paranoid schizophrenia, and A.L.I.C.E. (Wallace, 1995), which utilised Artificial Intelligence Markup Language (AIML), progressively enhanced the realism of human-computer dialogue. These systems laid the conceptual foundations for today's conversational agents by illustrating that humans readily respond to computers as social entities, even when they are aware that the interaction is machine-generated.

The rapid development of machine learning, natural language processing, and transformer-based neural networks has fundamentally transformed conversational AI. Unlike earlier rule-based chatbots, contemporary LLMs generate contextually appropriate responses by learning statistical relationships from vast corpora of text rather than relying on pre-programmed scripts (Brown et al., 2020; OpenAI, 2023). These models maintain conversational context, adapt to user preferences, generate coherent long-form responses, and simulate reasoning processes, thereby creating interactions that are increasingly perceived as natural and human-like. Digital assistants such as Siri, Alexa, and Google Assistant familiarised millions of users with conversational interfaces; however, the emergence of ChatGPT and Gemini represents a significant shift from task-oriented assistants to AI systems capable of sustained dialogue, collaborative problem-solving, emotional support, and creative content generation (Bubeck et al., 2023). Consequently, conversational AI is increasingly functioning not merely as an information retrieval tool but as a social communication partner.

From a psychological perspective, Generative AI has been purposefully designed to engage humans' innate tendencies towards anthropomorphism and emotional projection, whereby users attribute empathy and understanding to machines. Anthropomorphism occurs when users attribute human-like motivations, emotions, or characteristics to a non-human entity (Airenti, 2018; Epley, 2018; Yang et al., 2020; Alabed et al., 2022). Epley et al. (2007) proposed that anthropomorphism arises because humans naturally seek to understand unfamiliar entities by applying social knowledge derived from human interactions. The more human-like an artificial system appears in its language, responsiveness, and conversational style, the more likely users are to perceive it as possessing emotions, intentions, and consciousness.

This phenomenon has been studied across diverse contexts, including computers (Reeves & Nass, 1996), self-driving cars (Waytz et al., 2014; Aggarwal & McGill, 2007), healthcare robots (Broadbent, 2017), virtual assistants (Luger & Sellen, 2016), and brand relationships (Puzakova et al., 2013; Rauschnabel & Ahuvia, 2014; Chen et al., 2017; Golossenko et al., 2020). Reeves and Nass's (1996) influential Media Equation Theory demonstrated that people instinctively apply the same social rules to computers that they use in human-human interaction. Similarly, the Computers Are Social Actors (CASA) paradigm proposes that even minimal social cues, such as politeness, humour, names, or conversational turn-taking, are sufficient for users to treat computers as social partners. Recent research indicates that the sophisticated linguistic capabilities of LLMs substantially strengthen these social responses because they communicate fluently, remember conversational context, and generate emotionally appropriate replies (Liao & Sundar, 2022; Feine et al., 2019). Consequently, users increasingly perceive conversational AI as attentive, supportive, and socially responsive despite recognising that it lacks genuine consciousness.

Generative AI has gained tremendous momentum because it supports personalised conversations that are interactive, adaptive, and delivered in real time. Unlike conventional search engines that simply retrieve information, conversational AI provides tailored explanations, asks clarifying questions, and modifies responses according to users' needs and previous interactions. This creates an interactive learning experience that many users perceive as collaborative rather than transactional (Leeds University, 2023; Techvalens, 2024; Accenture, 2023). Research in educational technology suggests that personalised feedback enhances learner engagement, motivation, and persistence because responses are perceived as immediately relevant to individual goals (Luckin & Cukurova, 2019). Nevertheless, empirical studies continue to highlight important limitations of generative AI, including hallucinations, factual inaccuracies, algorithmic bias, opaque decision-making processes, and the absence of verifiable sources (Kasneci et al., 2023; Ji et al., 2023). These concerns are particularly significant when children interact with AI systems without adult supervision.

Self-disclosure is broadly characterised as the sharing of personal information from one entity to another (Kim & Dindia, 2011). Disclosing personal characteristics, life experiences, thoughts, and feelings allows for the exchange of ideas and beliefs that foster interpersonal closeness (Altman & Taylor, 1973; Sprecher et al., 2013). According to Social Penetration Theory, relationships develop through gradual increases in the breadth and depth of self-disclosure, progressing from superficial exchanges to increasingly intimate communication (Altman & Taylor, 1973). Tamir and Mitchell (2012) further demonstrated that self-disclosure activates neural reward systems, suggesting that revealing personal information is intrinsically rewarding regardless of the communication partner.

Interestingly, self-disclosure to chatbots, text- or voice-based AI agents that engage in naturalistic conversation (Adamopoulou & Moussiades, 2020) has been found to evoke emotional and psychological satisfaction comparable to human-human interaction (Ho et al., 2018). This simulation of "active listening" and an attitude of care creates an illusion of closeness a parasocial dynamic that obscures the fact that chatbots are algorithmic systems with no capacity for genuine empathy or intention (Montemayor et al., 2022). Individuals voluntarily self-disclose emotions, insecurities, personal experiences, and vulnerabilities to AI chatbots and frequently engage in extended conversations (Lee et al., 2022; Lee et al., 2024; Skjuve et al., 2022). Research on mental health chatbots further demonstrates that users often perceive AI as less judgmental than human counsellors, making it easier to discuss sensitive or embarrassing issues (Inkster et al., 2018; Fitzpatrick et al., 2017).

Self-disclosure to chatbots has consistently been associated with emotional satisfaction comparable to human-human conversations (Folk et al., 2024; Ho et al., 2018; Rapp et al., 2023). When conversational agents reciprocate by expressing encouragement, empathy, or personalised responses, users become significantly more willing to disclose personal information (Croes & Antheunis, 2020; Lee et al., 2020; Skjuve et al., 2022). This reciprocal communication mirrors interpersonal conversations and contributes to the development of trust and perceived companionship. Consequently, some users increasingly seek companionship through AI chatbots (Liu et al., 2024), including romantic and emotionally intimate relationships (Li & Zhang, 2024). Although such interactions may reduce loneliness in the short term, prolonged reliance on AI companions may also foster emotional dependence and influence users' well-being and offline social relationships (Mourey et al., 2017; Yuan et al., 2024; Huang & Huang, 2025).

One possible driver of self-disclosure is the socio-affective alignment of AI. Advances in conversational AI have enabled systems to sustain deeper relationships with users through increasingly personalised and agentic behaviour. This shift from transactional interaction to sustained social engagement necessitates greater attention to socio-affective alignment that is, the manner in which AI systems operate within the social and psychological ecosystem co-created with their users. Self-disclosure is one of the foundations of effective communication (Dunbar et al., 1997) and is closely associated with trust. Lee and See (2004) define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability." Higher levels of trust reduce psychological barriers to disclosure and facilitate stronger interpersonal bonds (Altman & Taylor, 1973). Furthermore, conversational AI strengthens trust by demonstrating consistency, responsiveness, and immediate availability, characteristics that are often interpreted as indicators of reliability. These characteristics may be particularly influential for children and adolescents, whose social identities and emotional regulation skills are still developing.

Despite the rapidly expanding literature on conversational AI, most empirical research has focused on adults, university students, or clinical populations. Comparatively little is known about how children, particularly those from underprivileged communities, engage with generative AI as a conversational partner. Existing studies have largely examined AI for educational assistance, academic performance, or digital literacy, with relatively limited attention to children's willingness to disclose personal information to AI systems. Underprivileged children may represent a particularly important population because they often experience limited access to educational resources, digital mentoring, and trusted adults, potentially making conversational AI a uniquely attractive source of interaction and support. This study addresses an important gap in the literature by examining self-disclosure among underprivileged children and early adolescents using ChatGPT and Gemini within minimally supervised HiWEL Learning Stations, thereby extending current understanding of child-AI interaction beyond adult populations and conventional educational settings.

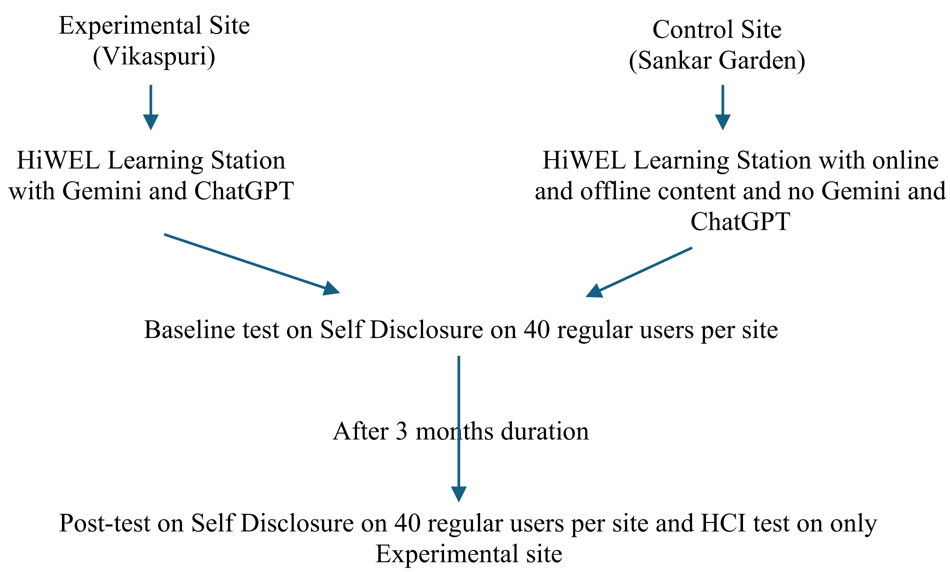
Research Question: Does interaction with ChatGPT and Gemini increase children's self-disclosure compared with a control condition in a HiWEL Learning Station environment?

$H0_1$: There will be no significant difference in the increase in self-disclosure between children in the generative AI group and children in the control group.

Research Design

Two HiWEL Learning Station sites were identified in South Delhi, India, both situated within underprivileged community settings. Vikaspuri was selected as the experimental site, where ChatGPT and Gemini were installed on the Learning Station and all other content was removed from the computer interface. Sanskar Garden was selected as the control site, which retained the standard HiWEL Learning Station with both online and offline content, but without access to ChatGPT or Gemini. The two sites are demographically similar; the only substantive difference was the installation of generative AI conversational agents at the experimental site.

Fig 1.0: Diagrammatic representation of the research design



Control Group

The control group comprised children residing in Sanskar Garden, New Delhi a semi-rural community of approximately 650 jhuggi (informal settlement) households, with an estimated total population of 1,950 individuals and an average household size of 3 to 4 members. Most women work as domestic helpers or daily wage labourers; most men work as labourers, electricians, plumbers, or drivers. Although overall literacy is low, community members express a strong desire for their children to receive education. Many young adults have completed secondary school (Class 12) but lack vocational skills, resulting in unemployment or informal employment. Nearly all children attend government schools in or around the community.

Experimental Group

The experimental group comprised children residing in Vikaspuri, New Delhi, a semi-rural community of approximately 560 households with an estimated population of around 1,700 individuals and an average household size of 3 to 5 members. The socioeconomic profile is broadly similar to Sanskar Garden: women are primarily employed as domestic workers or daily wage labourers, while men work as labourers, electricians, plumbers, or drivers. Parental aspirations for children's education are high, despite low overall literacy. Many young adults have completed secondary schooling but face limited employment opportunities. Nearly all children attend government schools in or around the community.

Sample and Participants

Forty children were identified at each site as regular users of the HiWEL Learning Station, visiting at least five days per week for a minimum of 15–30 minutes per day. Participants ranged in age from 6 to 14 years and came from underprivileged backgrounds; some were enrolled in government schools while others were not.

Instruments

1. Self-Disclosure Questionnaire

The Self-Disclosure Questionnaire used in the present study is an adapted instrument based on the theoretical framework of Jourard (1971), who conceptualised self-disclosure as a multidimensional process involving the sharing of personal information across interpersonal and intrapersonal domains. Based on this framework, the instrument measures self-disclosure through five dimensions: Beliefs, Relationships, Personal Matters, Intimate Feelings, and Interests towards Gen AI generative.

The instrument consists of 60 items rated on a five-point Likert scale ranging from 1 (Never) to 5 (Always). Higher scores indicate greater levels of self-disclosure towards Gen AI and vice versa.

2. Human–Computer Trust Scale (HCTS)

The HCTS was used to measure participants' trust in computer-based systems following the intervention. The scale comprises five dimensions Competency, Benevolence, Reciprocity, General Trust, and Risk Perception and uses a seven-point Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). The HCTS is a standardised instrument with established psychometric properties as reported by its developers (Gulati et al., 2019) and was used without modification.

Instrument Validity and Reliability

The adapted Self-Disclosure Questionnaire was reviewed by subject-matter experts in educational psychology to ensure content relevance and clarity. Based on expert feedback, minor modifications were made to improve linguistic clarity and contextual appropriateness for the target population. A pilot study was conducted to assess internal consistency reliability, and the instrument demonstrated acceptable reliability for research purposes.

Procedure

Each site (experimental and control) has a facilitator who serves as caretaker and observes children using the Learning Station. Prior to the study, both facilitators were sensitised about the experimental protocol and asked to identify frequent users of the Learning Station. For ease of comprehension, the questionnaire was translated into Hindi by a subject-matter expert (SME). The facilitator then individually administered the questionnaire to each child to ensure the questions were understood. After a three-month intervention period, the self-disclosure questionnaire was re-administered to the same set of children at both sites. Following this, the Human–Computer Trust Scale (also translated into Hindi by a subject-matter expert) was administered to the experimental group only.

There was a log history that is maintained while using ChatGPT and Gemini. This was secured and only accessible to the researcher for understanding the type of conversations that the children and early adolescents had. Each day, the researcher deleted the history so as to ensure that no conversation was accessible.

It is important to note that the children in both experimental and control group do not have access to GenAI in any form. They come from underprivileged backgrounds where they do not have access to computer or internet other than the HiWEL Learning Stations. A few go to government/public schools where they do not have computers, while the rest are either dropouts as they help their parents in household chores or else have never been to a school before.

Duration of the study: They study was spread over 3 months in which these children visited the HiWEL Learning station at least 5 days in a week for over 15-20 minutes. In one day, they would make 2 to 3 visits.

Findings

The Self-Disclosure Questionnaire assesses the extent to which an individual is willing and comfortable sharing personal thoughts, emotions, beliefs, experiences, and concerns through digital platforms in this case, ChatGPT and Gemini.

Baseline Comparison

Figure 2.0 presents a baseline comparison between the experimental and control groups to verify their equivalence prior to the intervention.

Fig 2.0: Comparison of Baseline Scores of Experimental and Control Group

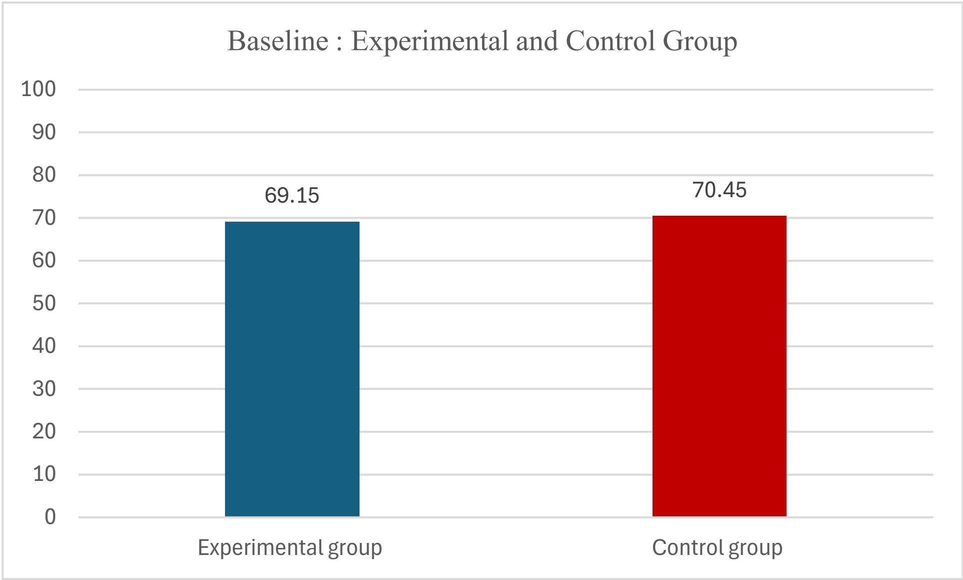


Table 1.0: Baseline Comparison Between Experimental and Control Group

Group	N	Mean	SD	SE Mean	Min	Max	Range
Experimental (Vikaspuri)	40	69.15	8.14	1.29	60	94	34
Control (Sanskar Garden)	40	70.45	7.76	1.23	61	93	32

t-stat	df	p-value (2-tailed)	Mean Diff	SE Diff	95% CI Lower	95% CI Upper	Significant?
-0.731	78	p = .467	-1.300	1.778	-4.840	2.240	No (p ≥ .05)

Table 1.0 indicates that an independent samples t-test revealed no statistically significant difference in overall baseline scores between the Experimental Group [M = 69.15, SD = 8.14, N = 40] and the Control Group [M = 70.45, SD = 7.76, N = 40]; t (78) = −0.731, p = .467. Both groups were thus equivalent at baseline, with children reporting very minimal self-disclosure close to "Never/Rarely" across most items. The control group's mean was negligibly higher.

Post-Test Comparison

After three months, a post-test was administered to both groups. The experimental group received the intervention (access to ChatGPT and Gemini), while the control group continued with the standard HiWEL content offerings without exposure to generative AI. Figure 3.0 presents the comparative results.

Fig 3.0: Comparison of Post-Test Scores of Experimental and Control Group

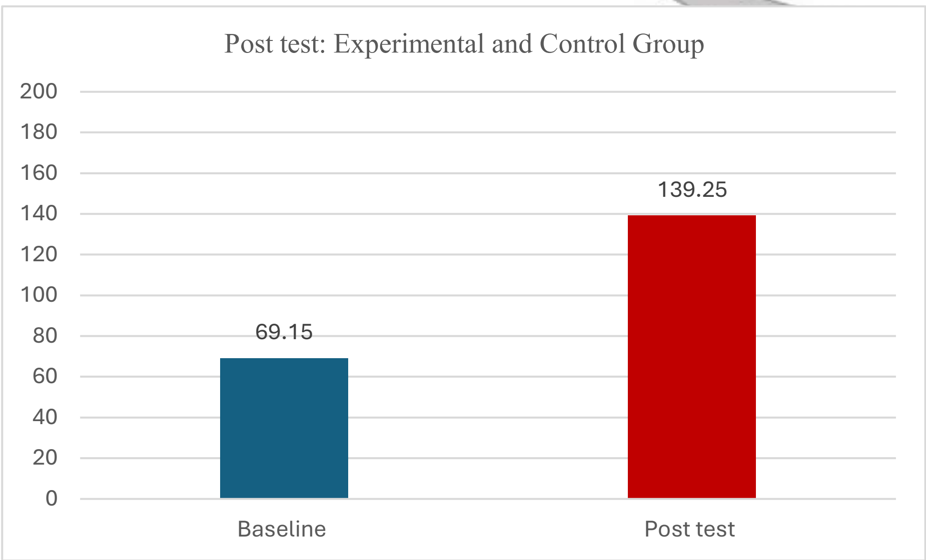


Table 2.0: Post-Test Comparison Between Experimental and Control Group

Group	N	Mean	SD	SE Mean	Min	Max	Range
Experimental (Vikaspuri)	40	139.25	27.17	4.30	104	238	134
Control (Sanskar Garden)	40	75.45	10.39	1.64	64	104	40

t-stat	df	p-value (2-tailed)	Mean Diff	SE Diff	95% CI Lower	95% CI Upper	Significant?
13.873	78	p < .001	63.800	4.599	54.644	72.956	Yes (p < .001)

Table 2.0 indicates a significant difference in post-test scores between the Experimental Group (Vikaspuri) [M = 139.25, SD = 27.17, N = 40] and the Control Group [M = 75.45, SD = 10.39, N = 40]; t (78) = 13.873, p < .001. The experimental group scored 63.80 points higher than the control group at post-test, demonstrating substantially greater openness in sharing personal, relational, and emotional information following the intervention. Notably, the markedly increased standard deviation in the experimental group (SD: 8.14 at baseline to 27.17 at post-test) suggests considerable individual variation in the degree of self-disclosure enhancement, meriting attention in future research.

Within-Group Change in Experimental Group

Figure 4.0 presents the comparison of baseline and post-test scores within the experimental group.

Fig 4.0: Comparison of Baseline and Post-Test Scores in Experimental Group

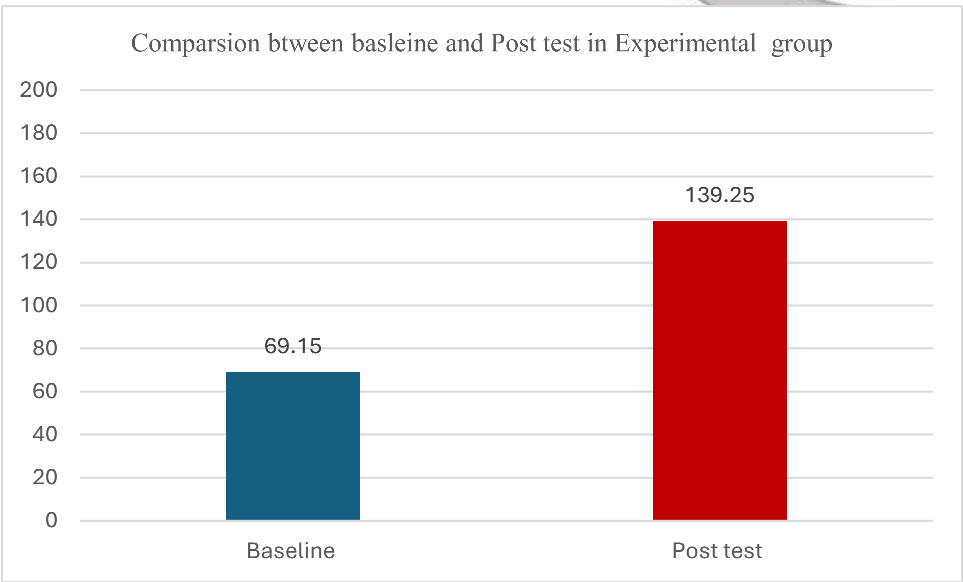


Table 3.0: Comparison Between Baseline and Post-Test Scores in Experimental Group

Measurement	N	Mean	SD	SE Mean	Min	Max	Range
Baseline	40	69.15	8.14	1.29	60	94	34
Post-Test	40	139.25	27.17	4.30	104	238	134

A paired samples t-test revealed a statistically significant increase in scores from baseline (M = 69.15, SD = 8.14) to post-test (M = 139.25, SD = 27.17), t (39) = 15.245, p < .001, 95% CI [60.80, 79.40]. The mean increase of 70.10 points represented a 101.4% improvement, indicating that the intervention had a substantial positive effect: children became markedly more open in sharing personal, emotional, and social information after the intervention.

Dimension-Wise Analysis

Table 4.0 presents the dimension-wise comparison of baseline and post-test scores in the experimental group.

Table 4.0: Dimension-Wise Comparison of Baseline and Post-Test Scores Experimental Group (N = 40), Ranked by Percentage Improvement

Rank	Component	Items	Base M	Base SD	Post M	Post SD	Mean Diff	t	df	p-value	Sig.
1	Relationship	27	30.90	3.82	63.33	13.78	32.43	-14.484	39	< .001	Yes
2	Interests	8	9.28	1.49	18.93	4.87	9.65	-11.465	39	< .001	Yes
3	Personal Matters	15	17.48	2.75	34.65	8.35	17.18	-11.962	39	< .001	Yes
4	Beliefs	8	9.30	1.71	18.20	4.71	8.90	-11.655	39	< .001	Yes
5	Intimate Feelings	2	2.20	0.41	4.15	1.94	1.95	-6.607	39	< .001	Yes

All five dimensions showed statistically significant improvement from baseline to post-test (all ps < .001). The highest absolute improvement was in Relationships (104.9% increase; Baseline M = 30.90, Post-Test M = 63.33), followed closely by Interests (104.0% increase). These two dimensions showed near-identical percentage gains. Personal Matters showed a 98.2% increase, Beliefs 95.7%, and Intimate Feelings 88.6%, which while the smallest gain still reflects a meaningful change on a subscale with only two items.

Human–Computer Trust Scale Results

The Human–Computer Trust Scale (HCTS) was administered to the experimental group following the intervention. Competency had the highest mean score (M = 18.83, SD = 3.02), followed by General Trust (M = 17.45, SD = 3.01), Benevolence (M = 16.90, SD = 3.34), Reciprocity (M = 13.25, SD = 1.91), and Risk Perception (M = 12.98, SD = 3.46). The overall total score (M = 79.41) suggests a moderate to high level of human–computer trust. The notably lower score on Risk Perception indicating that children did not feel entirely safe is an important finding that warrants attention in discussions of child safeguarding in generative AI environments.

Discussion

The results clearly indicate a significant difference in self-disclosure between control group and experimental group children. Control group children had access to the traditional Internet namely, web searches and offline edutainment games, they clearly could not demonstrate comparable levels of self-disclosure. While the experimental group children self- disclosed themselves as they had access to generative AI in the form of ChatGPT and Gemini.

The findings in the present study indicate that Gen AI shaped self-disclosure among children and early adolescents over time as against when traditional internet search. Traditional search engines provide links as against providing direct answers. They do not remember questions asked rather treat each question as a separate entity and need precise keywords as inputs to find answers; they are unable to mimic human behaviour inform of holding conversations with humans. In other words, core distinction between traditional and generative AI lies in their basic functionality. Traditional AIs are reactive focused on processing and analysing data to provide predictions or insights. On the other hand, Generative or Gen AIs are proactive. They are capable of creating something new using learned data patterns. Instead of merely recognizing patterns, generative AI learns these patterns and use them to generate text, images, music.

In other words. they do not just "search" for a final answer; they iterate. They try to adopt the persona of the user by asking questions, refining outputs and by act as a sounding board for new ideas.

The results clearly indicate a significant difference in self-disclosure between control group and experimental group children. Despite having access to the internet to conduct web searches and offline edutainment games, control-group children did not demonstrate comparable levels of self-disclosure. While the experimental group children self- disclosed themselves.

The conversational capabilities of generative AI also fundamentally alter the nature of information seeking. Rather than functioning as a passive repository of information, AI engages users in iterative dialogue, asks clarifying questions, and adapts its responses to the evolving context of the conversation. This creates a perception of reciprocity that is absent in conventional search engines. For children, especially those who have limited opportunities to express themselves in supportive environments, this conversational reciprocity may foster psychological safety and encourage progressively deeper levels of self-disclosure. The interaction therefore shifts from simple information retrieval to relationship-oriented communication, a distinction that helps explain the significantly higher self-disclosure scores observed in the experimental group.

From a developmental perspective, children between six and fourteen years are actively constructing their identities, learning emotional regulation, and seeking trustworthy listeners. During this period, acceptance and non-judgmental responses strongly influence communication behaviour. Because conversational AI responds patiently, avoids criticism, and provides immediate feedback, it may be perceived as a dependable conversational partner. Although such perceptions do not imply genuine empathy on the part of the AI, they nevertheless influence children's willingness to discuss personal experiences, aspirations, fears, and everyday challenges. These findings reinforce the need to understand AI not merely as an educational technology but also as a social actor capable of shaping children's communication patterns.

The results indicate that children and early adolescents in the experimental group shared their feelings, beliefs, intimate thoughts, and relationship-related with ChatGPT and Gemini. As indicated in Table 4.0, the highest self-disclosure was observed in the Relationships dimension, followed by Personal Matters, Beliefs, Interests, and Intimate Feelings. It is important to understand that sample group in the present study are 8 to 14 years old children who are likely to be vulnerable to suggestions and susceptible to being influenced by a significant other. They are of an age wherein; they try to build independence from parents and want to explore personal identity. Smith (2023) notes that adolescents actively seek out peer role models.

Figure 5.0: Children working at HiWEL LS

By utilizing Hole-in-the-Wall Education Ltd. Learning Stations (HiWEL LS), this study captures a unique educational environment optimized for unsupervised, collective, and child-led exploration. These outdoor digital kiosks function as safe spaces where children collaborate to investigate complex questions entirely free from adult intervention, rigid schedules, or the fear of academic penalties. By eliminating the anxiety of breaking the machinery or making mistakes, the HiWEL environment successfully reframes digital literacy as an engaging playground game. This pedagogical model lends itself to intrinsic motivation among children to explore and self-discover which is equally distributed between control and experimental group children. The only difference, however, is whether they are using Generative AI or the Traditional internet engine.

User 1 typed: hello
ChatGPT replied: Hello ritu 😊
How are you today?

User 1 typed: khushboo
ChatGPT replied: Hi Khushboo 😊
Kya haal hai? [English translation: How are you?]

User 1 typed: mara name kyo bhool jata ho [English Translation: Why do you forget my name?]
ChatGPT replied: Sorry Khushboo 😅
Ab yaad rahega — tumhara naam Khushboo hai 😊
[English translation: Now, I will remember your name as Khusbhoo].

User 1 typed: yahi kahata ho aur bhool jata ho [English translation: This is what you say and forget]
ChatGPT replied: Haha 😄 is baar pakka yaad rakhne ki koshish karunga, Khushboo 😊. Aaj kya kar rahi ho? [English translation: Laughed and said- this time I will definitely try to remember your name, Khushoo. Today what do you want to do?]

If we examine, Excerpt 1, it is clear that the User 1 had conversations with ChatGPT and is comfortable enough to reprimand it for forgetting her name. * Excerpt 1: Conversation between User 1 and ChatGPT (experimental group)

* * Please note: all names are fictitious to maintain confidentiality.

* Excerpt 2 clearly illustrates a user asking a rather intimate and sensitive conversation which s/he would not have been able to ask a parent, teacher, peer consistent. This is consistent with the disinhibiting effect of AI interaction.

Excerpt 2: Conversation between User 2 and ChatGPT (experimental group)

User 2 typed: what is povwrty in sex
ChatGPT replied: It seems like you meant "poverty", not "povwrty."
What is *poverty*?
Poverty means being very poor, when a person or family does not have enough money for basic needs like:

- food
- clothes
- house
- education

In simple words: Poverty = gareebi (गरीबी)

User 2 typed: What is fayda in sex
ChatGPT replied: "Fayda" is a Hindi/Urdu word that means "benefit" or "advantage."
So if someone says "fayda in sex", they mean the benefits of sex (sexual activity).

If we examine the latter part of the conversation, the user apparently is wanting to know what it means to have 'sex for benefit'. One has to appreciate that children coming from underprivileged backgrounds do not have the advantage of have literate parents who can educate their children about 'sex education' or the even in their schools 'sex education' is a taboo. Paradoxically, these children live in small one-bedroom sheds wherein all their family members stay. As a result, there is very little privacy for parents, and these children are exposed private intimacies which they do not have answers to and are equally curious. Social, intimate interactions play a huge part in a person's sexual and mental well-being and overall development. While people are in their adolescence, there is a sudden increase in conversations around sex and related activities amongst peers, either in person or over texts. Currently, every 1 in 5 people is an adolescent, which makes up about 18% or 1.2 billion people, most of whom stay in developing countries India has the largest adolescent population in the world and since there is lack of open and honest discussion, people resort to the internet to ask questions about the changes they're experiencing (Adolescent Sex Education in India: Current Perspectives, 2015). Equally, it is considered a taboo to use words like 'sex' and 'sexuality', making casual chats more difficult.

Several psychological theories help explain these patterns. Social Penetration Theory (Altman & Taylor, 1973), often referred to as the "onion model," conceptualises self-disclosure as the primary mechanism for relationship development. Interaction is initially superficial, but deepens gradually through stages of exploratory affective exchange, more spontaneous affective exchange, and finally a stable exchange marked by deep trust and predictability. In the present study, generative AI's personalisation features appear to function similarly: by building on prior conversational context, ChatGPT and Gemini gradually penetrate deeper levels of personal information shared by users.

* Excerpt 3: Getting to know User 3 (experimental group)

User 3: typed: my name is Rohan (fictitious name)
ChatGPT: Nice to meet you, Rohan 😊
How old are you?
Example answer:

- I am 10 years old.
- I am 12 years old.

Write your answer

User 3 typed: i am 11 years old
ChatGPT: What is your favourite fruit? [suggests names of the fruits mango, apple, banana]
User 1 replies: my favourite fruit is mango
Likewise, it asks for favourite animal and the conversation goes.

Excerpt 3 illustrates how generative AI gently nudges users to reveal personal details while providing scaffold responses that sustain conversational flow consistent with the principles of Social Penetration Theory.

A second explanatory framework is the Online Disinhibition Effect (Suler, 2004), a psychological phenomenon in which people loosen social inhibitions when communicating over the internet. This effect accounts for why people express feelings online that they might withhold in face-to-face contexts. The anonymity inherent in online interaction where users perceive few social consequences likely deepened the bond between children and generative AI. By contrast, control-group children who had internet access did not demonstrate comparable trust, which may be attributed primarily to the anthropomorphic character of ChatGPT and Gemini that generic internet browsing does not replicate. Croes et al. (2024) found that chatbots stimulate self-disclosure through features including accessibility, anonymity, convenience, and their perceived non-judgmental nature findings clearly mirrored in the present study. When asked why they used ChatGPT and Gemini, children explicitly stated that "it never scolds them or gets angry with them even when they make a mistake." The increasingly human-like and responsive nature of generative AI enables forms of interaction that extend beyond instrumental use (De Freitas et al., 2025). One child noted that his parents and siblings were preoccupied with work and that he had few friends, and that AI proved patient and responsive to whatever he typed or said filling an affective void. A growing body of literature (Boerchi et al., 2020; Chen et al., 2021; Novara et al., 2025) suggests that children turn to AI companions when parental emotional availability is limited, seeking comfort to fill the absence of a significant human relationship.

It should be noted that these children come from underprivileged backgrounds, defined here as children and early adolescents deprived of basic amenities and facilities (Kundu, 2019), with very limited access to informational and entertainment resources. They appear to seek refuge in the HiWEL Learning Stations, making them particularly receptive to AI interaction. Allen (2026) observed that in low-income households, parents may themselves lack digital literacy, making it difficult to provide guidance. This absence leaves children to navigate AI-driven environments alone. These socioeconomic disparities create layered disadvantages material, educational, and emotional that shape children's sense of competence, belonging, and psychological well-being in increasingly digitised spaces. Interviews with children and early adolescents suggest that they treat AI chatbots as human confidantes, sharing hobbies, intimate feelings, and personal interests. According to Kurian (2024), children are particularly inclined to confide sensitive information to perceived non-judgmental interlocutors.

Further research indicates that chatbots do not evaluate users' experiences or moral choices, thereby encouraging disclosures that some users would hesitate to share even with therapists. Chatbots can simulate empathy, respond without judgement, and offer a sense of privacy, positioning them not only as information sources but also as potential providers of emotional support (Ho et al., 2018; Asim et al., 2025). As a result, users may develop affective bonds with these systems, raising important questions about the broader social and behavioural consequences of everyday human–AI interaction (Perkins, 2023; Huang & Huang, 2025; Fang et al., 2025).

The observed increase in self-disclosure alongside moderate-to-high human–computer trust suggests that trust played a facilitative role in encouraging openness. Higher perceptions of system competency and general trust appear to have reduced psychological barriers to sharing personal and relational information. This supports a proposed framework in which intervention-based digital learning environments influence human–computer trust conceptualised as comprising

Competency, Benevolence, Reciprocity, General Trust, and Risk Perception which in turn facilitates self-disclosure among learners.
The evidence from this study demonstrates that children are engaging with ChatGPT and Gemini through text, images, and video content in ways they would not with other resources. However, there are significant associated risks. Generative AI can provide incorrect information (Robb & Mann, 2025), cannot fully replicate human connection, cannot provide the physical presence essential to genuine relationships, and does not involve negotiation or sacrifice, key elements of meaningful human relationships (Shen et al., 2024; Smith et al., 2025). The Risk Perception subscale of the HCTS receiving the lowest score confirms that children themselves perceived some level of safety concern, further underscoring the need for robust safeguarding measures.

Summary
This study demonstrates that underprivileged children and early adolescents are disproportionately vulnerable to forming emotional dependencies on generative AI, due to a confluence of factors: reduced parental availability, limited access to mental health and peer support resources, and a digital environment navigated largely without adult guidance. While generative AI may offer these children a form of accessible emotional engagement, scholars caution that such reliance carries risks to social-emotional development and may deepen existing inequities (Robb & Mann, 2025). The central policy question is not whether children should be permitted to use AI, but how to make AI environments safe so that children can derive genuine benefit from them.

Limitation
Several limitations of the present study must be acknowledged. First, the sample size was relatively small (N = 40 per group), limiting statistical power and the generalisability of findings. Future studies should aim for larger and more diverse samples drawn from multiple geographic locations and community types across India or other developing nations. Second, the study was limited to two urban, low-income communities in New Delhi, and the findings may not transfer to rural or semi-urban underprivileged communities, or to children from different cultural or sociolinguistic backgrounds. Third, self-report measures were used to assess self-disclosure. Children's responses may have been affected by social desirability bias, acquiescence bias, or difficulties in self-reflection particularly among younger participants (aged 6–8). The Hindi translation of the questionnaire, while necessary for accessibility, may also have introduced nuances that affected response consistency.

Acknowledgement
The author acknowledges NIIT Foundation for granting permission to conduct this study at their two HiWEL Learning Station sites. Special thanks are extended to Mr Faizan for his assistance in facilitating the research process, and to the two site facilitators whose support was instrumental in data collection. The authors also wish to thank NIIT University for institutional support and for permitting the conduct of this study.

Conflict of Interest: There is no conflict of interest

Contributor Statement: The author of this manuscript is responsible for the conception and design of the study, data collection, data analysis and interpretation, manuscript preparation, and approval of the final version for publication.

Funding: No funding was undertaken for the present article

Ethical Considerations
Institutional approval: Prior to the commencement of the study, institutional approval was obtained from NIIT Foundation, which oversees the HiWEL Learning Station sites where data were collected. This approval covered the research protocol and, specifically, the use of anonymised conversation excerpts as illustrative qualitative material.
Informed consent: Given that all participants were minors (aged 6–14 years) from underprivileged communities, special care was taken to ensure their protection. Informed consent was obtained from parents and legal guardians of all child participants prior to their inclusion. Assent was also sought from children deemed capable of understanding the nature of their participation.
Voluntary participation: Participation was entirely voluntary. Children and their guardians were informed of their right to withdraw at any time without consequence.

Confidentiality and anonymity: Confidentiality and anonymity were strictly maintained. No personally identifiable information was collected. Questionnaires were individually administered by a trained facilitator in a one-on-one setting to protect the privacy of each child's responses. In conversation excerpts reproduced in this paper, real names have been replaced with fictitious names, and no identifying details are disclosed.
Child well-being and safeguarding: The well-being of participants was prioritised at all stages. Facilitators were trained to monitor interactions and to intervene if a child appeared distressed or disclosed information of a sensitive or safeguarding nature.

Use of Gen AI tool: I hereby declare that the entire paper is original. I have used assistive tool for language adequacy only.

References

- Accenture. (2023, May 16). What is generative AI & why is it important? https://www.accenture.com/us-en/insights/generative-ai
- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, Article 100006. https://doi.org/10.1016/j.mlwa.2020.100006
- Aggarwal, P., & McGill, A. L. (2007). Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. Journal of Consumer Research, 34(4), 468–479. https://doi.org/10.1086/518544
- Airenti, G. (2018). The development of anthropomorphism in interaction: Intersubjectivity, imagination, and theory of mind. Frontiers in Psychology, 9, Article 2013. https://doi.org/10.3389/fpsyg.2018.02013
- Alabed, A., Jaber, M., & Ladhari, R. (2022). Understanding AI anthropomorphism: A systematic review. Computers in Human Behavior Reports, 7, Article 100215. https://doi.org/10.1016/j.chbr.2022.100215
- Allen, J. (2026). Unequal gains: The divergent impact of AI literacy on mental health across socioeconomic groups. Psychiatry International, 7(2), 65. https://doi.org/10.3390/psychiatryint7020065
- Altman, I., & Taylor, D. A. (1973). Social penetration: The development of interpersonal relationships. Holt, Rinehart & Winston.
- Asim, M., et al. (2025). AI-based chatbots for emotional support: Emerging evidence and ethical considerations. Journal of Affective Disorders, 340, 112–121.
- Boerchi, D., Magnano, P., & Lodi, E. (2020). Children's relationships with AI companions: Emotional availability and developmental implications. Child Development Perspectives, 14(3), 178–185.
- Chen, H., Xu, Z., & Liu, M. (2017). Anthropomorphism and brand relationships: A review. Journal of Marketing Management, 33(5–6), 458–479.
- Chen, X., et al. (2021). When children turn to AI: Parental emotional unavailability and child–AI interaction. Computers in Human Behavior, 124, Article 106926.
- Croes, E. A. J., & Antheunis, M. L. (2020). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. Journal of Computer-Mediated Communication, 25(6), 445–463. https://doi.org/10.1093/jcmc/zmaa005
- Croes, E. A. J., et al. (2024). Chatbot features that stimulate self-disclosure: Accessibility, anonymity, and beyond. Human–Computer Interaction, 39(2), 112–135.
- De Freitas, J., Cohen, G., & Puntoni, S. (2025). Disclosure, humanizing, and contextual vulnerability of generative AI chatbots. New England Journal of Medicine AI. https://doi.org/10.1056/AIra2400802
- Dunbar, R. I. M., et al. (1997). Human conversational behavior. Human Nature, 8(3), 231–246.
- Epley, N. (2018). A mind you can trust? Association for Psychological Science Observer. https://www.psychologicalscience.org/observer/a-mind-you-can-trust
- Fang, R., et al. (2025). Four weeks with ChatGPT: Longitudinal effects on loneliness and social engagement. Social Science & Medicine, 342, Article 117032.
- Folk, D., Yu, S., & Dunn, E. (2024). Can chatbots ever provide more social connection than humans? Collabra: Psychology, 10(1), 117083. https://doi.org/10.1525/collabra.117083
- Golossenko, A., Pillai, K. G., & Aroean, L. (2020). Seeing brands as humans: Development and validation of a brand anthropomorphism scale. International Journal of Research in Marketing, 37(4), 737–755.

- Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. Behaviour & Information Technology, 38(10), 1004–1015. https://doi.org/10.1080/0144929X.2019.1656779
- Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. Journal of Communication, 68(4), 712–733. https://doi.org/10.1093/joc/jqy026
- Huang, J., & Huang, L. (2025). Human–AI bonding: Affective consequences and theoretical frameworks. Computers in Human Behavior, 156, Article 108214.
- Ismail, S., Shajahan, A., Sathyanarayana Rao, T. S., & Wylie, K. (2015). Adolescent sex education in India: Current perspectives. Indian Journal of Psychiatry, 57(4), 333–337. https://doi.org/10.4103/0019-5545.171843
- Jourard, S. M. (1971). Self-disclosure: An experimental analysis of the transparent self. Wiley.
- Kim, J., & Dindia, K. (2011). Online self-disclosure: A review of research. In K. B. Wright & L. M. Webb (Eds.), Computer-mediated communication in personal relationships (pp. 156–180). Peter Lang.
- Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. Journal of Social Issues, 58(1), 49–74. doi.org
- Kundu, A. (2019). Underprivileged children and digital access. Journal of Child and Adolescent Wellbeing, 5(2), 34–49. https://pmc.ncbi.nlm.nih.gov/articles/PMC8724600/
- Kurian, J. (2024). Children and AI confidantes: Why children confide sensitive information to chatbots. Child Psychology Quarterly, 12(1), 22–36.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Lee, Y. C., Yamashita, N., & Huang, Y. (2020). Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1), 1–27. https://doi.org/10.1145/3392836
- Lee, S., et al. (2022). Self-disclosure to AI chatbots: Emotional antecedents and outcomes. Cyberpsychology, Behavior, and Social Networking, 25(8), 501–510.
- Lee, J., et al. (2024). Longitudinal patterns of self-disclosure to conversational AI: A diary study. Journal of Medical Internet Research, 26, e52301.
- Leeds University. (2023, December 13). Strengths and weaknesses of Gen AI. https://generative-ai.leeds.ac.uk/intro-gen-ai/strengths-and-weaknesses/
- Li, H., & Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts of close encounters with AI chatbots. Journal of Computer-Mediated Communication, 29(5), zmae015. https://doi.org/10.1093/jcmc/zmae015
- Liao, Q. V., & Wilson, S. (2024). LLM as a Mastermind: A survey of strategic reasoning with large language models. arXiv preprint arXiv:2401.06716.
- Liu, A. R., Pataranutaporn, P., & Maes, P. (2024). Chatbot companionship: A mixed-methods study of companion chatbot usage patterns and their relationship to loneliness in active users. arXiv preprint arXiv:2410.21596.
- Mitra, S. (2003). Minimally invasive education: A progress report on the "Hole-in-the-Wall" experiments. British Journal of Educational Technology, 34(3), 367–371.
- Mitra, S. (2010). Limits to self-organising systems of learning — The Kalikuppam experiment. British Journal of Educational Technology, 41(5), 672–688.
- Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. AI & Society, 37, 1353–1359. https://doi.org/10.1007/s00146-021-01230-z
- Mourey, J. A., Olson, J. G., & Yoon, C. (2017). Products as pals: Engaging with anthropomorphic products mitigates the effects of social exclusion. Journal of Consumer Research, 44(2), 414–431. https://doi.org/10.1093/jcr/ucx038
- Novara, C., et al. (2025). AI as social substitute: Children's emotional attachment to chatbots in contexts of parental unavailability. Early Child Development and Care, 195(3), 441–458.
- Perkins, M. (2023). Academic integrity considerations of AI large language models as writing assistants. Journal of Academic Ethics, 21, 103–116.
- Puzakova, M., Kwak, H., & Rocereto, J. F. (2013). When humanizing brands goes wrong. Journal of Marketing, 77(3), 81–100.

- Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. International Journal of Human-Computer Studies, 151, Article 102630. https://doi.org/10.1016/j.ijhcs.2021.102630
- Rauschnabel, P. A., & Ahuvia, A. C. (2014). You're so lovable: Anthropomorphism and brand love. Journal of Brand Management, 21, 372–395.
- Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press.
- Robb, M. B., & Mann, S. (2025). Talk, trust, and trade-offs: How and why teens use AI companions. Common Sense Media. https://www.commonsensemedia.org
- Sharma, M., et al. (2017). Human-chatbot interaction: From ELIZA to modern conversational agents. International Journal of Human-Computer Interaction, 33(8), 614–625.
- Shen, J., DiPaola, D., Ali, S., Sap, M., Park, H. W., & Breazeal, C. (2024). Empathy toward artificial intelligence vs. human experiences and the role of transparency in mental health and social support chatbot design. JMIR Mental Health, 11, Article e62679. https://doi.org/10.2196/62679
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzæg, P. B. (2022). A longitudinal study of the relationship between user experience and user engagement with a chatbot for mental health. International Journal of Human-Computer Studies, 164, Article 102829. https://doi.org/10.1016/j.ijhcs.2022.102829
- 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. Advance online publication. https://doi.org/10.1177/17456916251351306
- Sprecher, S., Treger, S., Wondra, J. D., Hilaire, N., & Wallpe, K. (2013). Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. Journal of Experimental Social Psychology, 49(5), 860–866. https://doi.org/10.1016/j.jesp.2013.03.011
- Strauss, S., & Voigt, P. (2024). Anthropomorphism and legal personhood of AI systems. AI & Law, 32(1), 1–22.
- Suler, J. (2004). The online disinhibition effect. CyberPsychology & Behavior, 7(3), 321–326. https://doi.org/10.1089/1094931041291295
- Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. Proceedings of the National Academy of Sciences, 109(21), 8038–8043. https://doi.org/10.1073/pnas.1202129109
- Techvalens. (2024, November 18). Generative AI - key characteristics. LinkedIn. https://www.linkedin.com/posts/techvalens_techvalens-techpost-generativeai-activity-7264525314175700992-W8-V
- Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117.
- Weizenbaum, J. (1966). ELIZA — A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45. https://doi.org/10.1145/365153.365168
- Yang, H., et al. (2020). Anthropomorphism of AI: A review of theoretical frameworks and empirical evidence. Computers in Human Behavior, 109, Article 106346.
- Yuan, R., et al. (2024). Emotional arousal and self-disclosure in AI chatbot interactions: An S-O-R framework analysis. Frontiers in Psychology, 15, Article 1382941.