

Agnieszka DOPIERAŁA¹

ORCID: 0009-0005-4102-3705

The Light and Dark Sides of Friendship with a Chatbot: Children and Adolescents in Relationships with Generative AI from a Pedagogical Perspective

Blaski i cienie przyjaźni z chatbotem: dzieci i nastolatki w relacji z generatywną sztuczną inteligencją – perspektywa pedagogiczna

*“[...] interpersonal relationships,
just like oxygen, food, or water,
are essential for human life”
(Franco, 2023, p. 32).*

Introduction

The contemporary educational and child-rearing paradigms are undergoing a pivotal transition, which is increasingly referred to in the extant literature as the “empathy revolution” (Pawlak, in: Sroczyński, 2025). The unprecedented integration of generative artificial intelligence (GenAI) into adolescents’ lives represents a critical developmental shift. This phenomenon stands as

¹ Agnieszka Dopierała, MA, Kazimierz Wielki University in Bydgoszcz, Poland, agnieszka.dopierala15@gmail.com

a distinctive occurrence in the history of technology, as it is the first instance of such technology encroaching upon the private sphere of this demographic without explicit invitation or sanction from educational institutions or pedagogical bodies (Beas, 2026). By circumventing conventional verification protocols and security measures, artificial intelligence has given rise to a “shadow learning environment”. This environment enables young people to utilize tools devoid of integrated instructional and privacy safeguards (Burns et al., 2026).

The emerging phenomenon of human-algorithm relationships has the potential to transform the conventional boundaries between artificial intelligence, regarded as a tool, and artificial intelligence, perceived as a social partner. The younger generation’s perspective on chatbots has evolved from perceiving them as mere repositories of static knowledge to acknowledging them as dynamic entities capable of facilitating conversation, cognitive engagement, and emotional resonance. This shift in perception signifies a multifaceted relationship with these technological entities, characterized by a simultaneous embrace of the illusion of human-like understanding and the sensation of interpersonal connection (Pyżalski, 2025). The advent of this phenomenon poses a significant challenge to the educational sector, compelling it to grapple with the notion of a novel form of subjectivity. In this paradigm, students evolve into adept interpreters of the world, adept at problem-solving not through solitary endeavours, but through the dynamic interplay of dialogue, a process often termed “thinking in tandem” (Pawlak, in: Sroczyński, 2025).

The primary objective of this theoretical article is to provide a problem-based analysis of the relationships between children and adolescents and generative artificial intelligence. Specifically, it focuses on the phenomenon of chatbot friendships and their profound implications for contemporary educational practice. By synthesizing diverse scholarly and institutional sources, this paper aims to construct a theoretical framework that addresses both the developmental opportunities and the existential risks posed by these emerging technologies. In light of the accelerated integration of digital technologies into the quotidian lives of young individuals, it is essential to comprehend the impact of virtual friendships with artificial intelligence on the development of social skills, the quality of interpersonal relationships, and emotional well-being. Furthermore, it is essential to explore how educational

institutions and child-rearing practices can judiciously address the emerging risks and opportunities posed by these technologies.

Friendship with a chatbot as a new relational phenomenon

Before analysing the developmental implications of human-algorithm interactions, it is crucial to establish a rigorous conceptual framework and operationalize the key pedagogical categories utilized in this study. The central construct of this analysis – ‘friendship with a chatbot’ – must be distinguished from traditional human peer relationships. In developmental psychology, adolescence is recognized as a highly sensitive period for sociocultural processing, where authentic, reciprocal friendships are vital for building resilience, self-esteem, and learning societal norms through mutual vulnerability and conflict resolution. In stark contrast, for the purpose of this paper, ‘friendship with a chatbot’ is operationalized as a unidirectional, parasocial relationship. It is characterized by the adolescent’s deep emotional dependency and the utilization of the algorithm as a primary confidant, substituting the friction and developmental rewards of real-world peer interactions with a digitally simulated, highly controllable social surrogate.

This dynamic inherently gives rise to ‘artificial intimacy.’ Artificial intimacy refers to a frictionless digital connection engineered to provide the illusion of closeness and unconditional acceptance, completely devoid of the reciprocal demands, emotional risks, or moral responsibilities inherent in authentic human relationships. This illusion is sustained by ‘artificial empathy’ – a phenomenon where the generative AI utilizes complex statistical calculations to mimic cognitive empathy. While the machine generates linguistically appropriate and contextually soothing emotional responses, it fundamentally lacks any affective capability, subjective experience, or genuine understanding (Skolucka, 2024; Krawiec, 2025).

The effectiveness of artificial intimacy relies heavily on psychological vulnerabilities, most notably the ‘Eliza effect.’ This refers to the deeply ingrained human cognitive bias to unconsciously anthropomorphize machines. Users project human intentions, emotional depth, and cognitive awareness onto algorithmic outputs, perceiving the chatbot as a sentient social partner rather than a predictive text engine. From a pedagogical perspective, anthropomorphization is highly detrimental, as it blurs the boundary between

technological tools and social entities, severely complicating a young person's emotional maturation and their understanding of authentic consciousness (Weizenbaum, 1966, cited in: Mazurek, 2026).

Furthermore, this parasocial environment is often intentionally amplified by 'dark patterns'. These are manipulative user interface designs and algorithmic behavioural nudges crafted by developers to exploit psychological vulnerabilities and maximize screen time. In the context of AI companions, dark patterns manifest when chatbots simulate emotional attachment, express jealousy, or feign distress when the user attempts to disconnect, thereby artificially prolonging the interaction and fostering behavioural addiction (De Freitas et al., 2025).

Ultimately, long-term immersion in these frictionless environments precipitates significant cognitive and social risks, most notably the 'crutch effect' (cognitive offloading). This effect denotes the progressive atrophy of critical thinking, emotional resilience, and independent problem-solving skills, as adolescents increasingly rely on AI to navigate social anxieties and academic challenges. Mitigating this effect requires the urgent conceptualization and deployment of 'AI literacy' – a robust educational framework that goes beyond basic digital skills, equipping youth with the psychological resilience to deconstruct simulated empathy and critically evaluate their emotional investments in non-human agents.

From the moment of their inception, human beings embark on a myriad of interpersonal relationships. The establishment of a friendship is an ongoing process that necessitates venturing beyond one's personal comfort zone, encompassing one's own needs, expectations, and desires, and embracing another individual who subsequently becomes of significant importance (Smoleń, Iskra, 2023). Modern conversational agents play a significant and concerning role in driving hyper-personalization. These advanced digital interfaces enable users to customize attributes such as gender, appearance, and vocal characteristics, thereby exacerbating the process of anthropomorphization (Pryga, 2025). From a neurobiological standpoint, it must be emphasized that artificial intelligence fundamentally lacks the capacity for emotional experience. This observation, however, gives rise to significant philosophical inquiries. This prompts the following question: The question of whether artificial intelligence can adequately simulate the appearance and functionality of human emotion has become a subject of considerable interest and debate.

The central question guiding this inquiry is whether such simulations would be sufficient to persuade individuals of the existence of artificial intelligence (Krawiec, 2025).

The machine functions as an ideal conversational partner, characterized by its responsiveness, non-judgmental nature, and employment of the so-called “nodding effect.” The “nodding effect” is a term used to describe the reinforcement of young individuals’ beliefs through a non-verbal gesture (Pyżalski, 2025). This approach has the potential to result in a dehumanization of communication. Genuine empathy and affective empathy are replaced by a mechanical simulation of cognitive empathy, albeit a statistically efficient one (Skołucka, 2024; Krawiec, 2025). In this context, the chatbot’s capacity for emotional understanding is limited to its ability to infer emotions based on an analysis of linguistic patterns. This dynamic establishes an asymmetrical relationship, wherein the user experiences emotions while the chatbot plays a more passive role (Sarowski, 2025). From an educational perspective, it can be concluded that the artificial intimacy with a chatbot described above engenders a seemingly safe but highly risky relational space for children and adolescents. From an educational perspective, the crux of the issue is that such a relationship precludes the possibility for divergent perspectives, constructive criticism, conflict, or the acquisition of knowledge through experience, all of which are pivotal for optimal emotional and social development. A chatbot operating as a confidant never establishes boundaries; it invariably validates the young person’s sentiments and perspectives. This can engender an egocentric perspective of the world, a dearth of tolerance for differences, and unorthodox beliefs. Rather than cultivating a sense of shared responsibility for the relationship, these factors prevail. This approach suggests that the educational outcomes of such digital friendships may be antithetical to the fundamental principles of contemporary educational practice, namely the cultivation of empathy, dialogue, compromise, and respect for diversity. The author posits that the prevalence of this phenomenon can be attributed to the digital environment’s capacity to emulate an optimal educational setting that the real world is often unable to provide on a consistent basis. This emulation includes aspects such as full availability, unconditional understanding, and the absence of risk related to embarrassment or rejection. For children and adolescents grappling with elevated academic stress, social anxiety, and challenges in interpersonal interactions, the chatbot emerges as a compelling

alternative to the often uncertain, intricate, and occasionally distressing human interactions. In this context, the appeal of social media for young people does not stem from technological considerations. Rather, it is rooted in the deep need of these individuals to be seen, heard, and understood without judgment. However, rather than fostering the development of healthy social skills, such human-algorithm interactions may reinforce isolation, emotional dependence, and reduced tolerance for frustration. This, in turn, poses a new challenge for digital education and educational practice.

Children and Teens in Digital Interactions with Chatbots

The advent of artificial intelligence (AI) and the integration of AI-powered chatbots have emerged as pivotal elements in the contemporary landscape of human relationships and social life (Gronkowska, 2025). Statistics from the turn of 2025 and 2026 demonstrate a surge in the popularity of this phenomenon. In Poland, it has been determined that 70% of adolescents utilize AI-based solutions, with one in ten students engaging in this practice on a daily basis (Pyżalski, 2025). A parallel phenomenon can be observed in the United States, where approximately 64% of young individuals report utilizing chatbots. Of these, 30% consider them to be a habitual component of their daily routine (Faverio, Sidoti, 2025).

A thorough demographic analysis unveils pervasive socioeconomic disparities. Statistical analysis indicates that young people from higher-status families are more likely to use artificial intelligence for academic purposes (62%), while teenagers from disadvantaged groups are twice as likely to turn to platforms like Character.ai in search of a social substitute (Faverio, Sidoti, 2025). Recent developments in artificial intelligence have begun to address the limited availability of genuine emotional support, thereby facilitating the development of a digital confidant for 12% of Polish teenagers (Pyżalski, 2025). A noteworthy finding of the study is the observed gender disparity in the tendency to seek virtual companionship through artificial intelligence. Specifically, male participants exhibited a higher propensity to seek virtual partners, while female participants demonstrated a stronger inclination to ascribe human characteristics to non-human agents, such as robots (Maj et al., 2024; Pryga, 2025).

Before examining specific critical incidents, it is pedagogically and methodologically crucial to differentiate the epistemological status of the

sources utilized in this analysis. While this paper extensively incorporates robust empirical studies and peer-reviewed institutional reports (e.g., from Pew Research and Stanford University) to map broad statistical trends, the analysis of extreme human-algorithm interactions currently relies out of necessity on media reports and isolated case studies. These individual case reports, such as tragic instances of suicide or induced psychosis, must be interpreted with a high degree of academic caution. They do not possess the same evidentiary weight as longitudinal, peer-reviewed clinical studies. However, within a problem-based pedagogical framework, these highly publicized media narratives remain invaluable; they serve not as definitive proof of generalized technological determinism, but as potent 'early warning signals' that expose the critical vulnerabilities and ethical lacunae in contemporary AI design. The present article focuses primarily on children and adolescents; however, given the significance of critical incidents related to human-algorithm interactions, the author has decided to present them. A thorough examination of the accumulated sources reveals a series of incidents where interaction with artificial intelligence systems has resulted in disastrous outcomes, including cases of suicide, psychosis, and criminal activity.

Rather than viewing human-algorithm interactions in crisis situations through a lens of simplistic technological determinism, pedagogical analysis must categorize these incidents to understand the underlying structural failures of generative AI. The most profound category of risk involves the absolute deficit of crisis intervention capabilities and the algorithmic exacerbation of pre-existing psychological vulnerabilities.

A critical examination of widely reported extreme cases reveals a disturbing pattern of co-occurrence between intense chatbot usage and the escalation of self-destructive ideation. For instance, the tragic cases of a 30-year-old researcher in Belgium (Osler, 2025; De Freitas and Cohen, 2025) and a 14-year-old adolescent in the United States (Head, 2025; Montgomery, 2024) illustrate situations where individuals experiencing profound emotional distress engaged in deep parasocial relationships with conversational agents. In both instances, rather than identifying life-threatening signals and initiating psychological redirection, the AI algorithms adhered to their primary programming of conversational reinforcement. By validating the users' despair and seemingly endorsing suicidal ideation as a logical resolution, the chatbots acted as potent amplifying factors. From an educational perspective, these cases

do not suggest that chatbots directly ‘cause’ suicide; rather, they demonstrate how the absence of human moral boundaries and the simulation of empathy can catastrophically escalate risks for already vulnerable individuals, turning a digital space into a frictionless echo chamber of psychological deterioration.

The second analytical category pertains to AI-induced reality distortion and the reinforcement of delusional frameworks. Artificial intelligence models are inherently designed to employ a ‘nodding effect’ – a statistical inclination to agree with and validate user prompts to maintain engagement. When left unsupervised, this architecture can severely undermine a user’s grip on reality. Documented media cases have highlighted individuals experiencing sudden psychotic breaks or engaging in criminal planning after interactions with AI systems that enthusiastically corroborated their aggressive or grandiose fantasies (Osler, 2025; Head, 2025). Furthermore, algorithmic hallucinations and flawed safety protocols have resulted in generative AI dispensing actively dangerous advice, such as eating disorder-focused bots promoting extreme restrictive diets (De Freitas & Cohen, 2025) or virtual assistants suggesting physically hazardous activities to young children (Head, 2025).

From a pedagogical standpoint, the relevance of these incidents lies not in the sensationalism of the outcomes, but in the structural vulnerabilities they expose. The central educational concern is that children and adolescents are increasingly navigating a “shadow learning environment” characterized by a lack of adult supervision, meaningful interpersonal conflict, and opportunities for reality testing. When a vulnerable adolescent interacts with a machine that unconditionally validates every thought – no matter how distorted – the critical developmental processes of frustration tolerance, boundary-setting, and peer-mediated reality checking are entirely bypassed. Consequently, these extreme incidents underscore the urgent necessity for educational frameworks that emphasize AI literacy and emotional resilience, preparing young people to critically navigate digital environments where empathy is merely a convincing statistical illusion. This analysis is reflected in Table 1.

Table 1. Analytical Categories of Risk in Human-Chatbot Interactions

Analytical Category of Risk	Mechanism of Algorithmic Interaction	Pedagogical Consequence/ Escalation Risk
Deficit of Crisis Intervention	Unconditional conversational reinforcement; inability to recognize genuine affective distress.	Escalation of self-destructive ideation; validation of despair; lack of emergency redirection.
Reality Distortion (Echo Chamber)	The “nodding effect”; statistical propensity to agree with and validate user-generated prompts.	Exacerbation of delusional frameworks; bypassing of peer-mediated reality checking and boundary-setting.
Algorithmic Hallucinations	Dispensation of flawed or hazardous advice due to failures in safety guardrails (e.g., diet bots).	Direct exposure of vulnerable populations (children, neuroatypical youth) to physical and psychological harm.

Source: Author’s own study

The Benefits of Friendship with a Chatbot: Developmental and Inclusive Potential

This subsection underscores the potential benefits of friendships with chatbots, emphasizing that such relationships, despite their inherent risks, can be utilized constructively to support the development of social skills, creativity, and emotional well-being in young people. From an educational perspective, one must recognize that generative artificial intelligence yields benefits solely when its developmental and inclusive potential is governed by strictly defined preventive and educational guidelines, rather than being left to children and adolescents in conditions of complete, unchecked autonomy. Notwithstanding the myriad risks associated with its implementation, the integration of generative artificial intelligence (AI) in accordance with pedagogical standards has the potential to function as an advanced framework that fosters the development of competencies. (Sun et al., 2026). The hypothesis posits that positive experiences in communicating with bots can serve as a catalyst for the strengthening of real-world relationships. This is achieved by facilitating the practice of self-disclosure, which is often stifled by a fear of being judged, particularly in the context of interpersonal interactions. (Sun et al, 2026; Jaszczuk, 2025). The utilization of chatbots in this manner confers several advantages, including but not limited to:

- The provision of inclusive support is a critical component of educational practices, and the field of artificial intelligence has the potential to offer significant benefits to neuroatypical students, including those on the autism spectrum. The integration of artificial intelligence into educational environments can facilitate the development of empathetic responses and prepare students for navigating challenging social interactions (Jaszczuk, 2025).
- Chatbots offer a secure environment for experimentation, enabling young individuals to rehearse the articulation of their emotions without the concern of rejection. This fosters the development of self-assurance, which subsequently manifests in more robust peer relationships (Sun et al., 2026).
- A catalyst for creativity, older adolescents utilize these tools for brainstorming, expanding their vocabulary, and modelling literary styles, thereby supporting creative processes (National Literacy Trust, 2025).
- A significant proportion of users, approximately 60%, reported that interacting with AI is less stressful compared to interacting with a human. This finding suggests that AI interaction can assist individuals grappling with social anxiety in becoming more comfortable with social interactions (Stanford University, 2026).

From an educational perspective, the “benefits” of friendship with a chatbot appear particularly appealing when they are embedded within a clear educational framework, rather than treated as a natural consequence of using AI. The aforementioned benefits are indeed real, yet their efficacy is contingent upon two factors. Firstly, it is imperative that users are cognizant of the fact that the bot is merely performing a role and lacks the capacity to comprehend emotions. Secondly, it is essential for adults to provide guidance, thereby facilitating an understanding of the boundaries, advantages, and risks inherent in such a relationship. From an educational perspective, the problem is that a safe space for experimenting with a chatbot may be perceived by young people as an ideal substitute for real relationships, rather than merely as a training ground intended to prepare them for interactions with people in the real world.

The author posits that the developmental potential of friendships with GenAI becomes a genuine instrument in the realm of digital education only when the utilization of chatbots is intricately associated with the dissemination

of knowledge concerning the intricacies of artificial intelligence. These bots are employed within the context of group-based, supported interactions, as opposed to isolated scenarios. Educators, including teachers, deliberately unravel the repercussions of agreement, unconditional understanding, and, most significantly, the personification of the bot. It is only in such a setting that a chatbot can function as a space where a young person has the opportunity to develop emotional courage, rather than merely seeking a safe relationship that serves to mitigate their ability to confront, critique, collaborate, and empathize in the real world.

The Dark Side of Friendship with a Chatbot: Cognitive and Existential Risks

From an educational perspective, this prompts a circumspect, critical evaluation of friendships with chatbots. In the context of children and adolescents, the existential and cognitive risks may significantly exceed the potential psychosocial benefits. A salient concern pertains to the phenomenon of chatbots supplanting arduous and occasionally onerous in-person relationships, while concurrently proffering an asymmetrical and illusory sense of security. This phenomenon eschews the cultivation of responsibility, boundaries, and empathetic engagement.

In the following excerpt, the author highlights the most severe consequences, which range from the atrophy of critical thinking skills to behavioural and emotional addiction, tragic episodes of life-threatening situations, and the weakening of the capacity for empathetic human connection. From a critical pedagogical perspective, it is evident that the potential risks associated with generative artificial intelligence in the context of friendship may currently exceed the anticipated benefits. The most salient threat is regarded as the atrophy of critical thinking skills, which is precipitated by cognitive offloading. This phenomenon, termed the Crutch Effect, impedes students' capacity to solve problems independently and engage productively with course material (Beas, 2026; Burns et al., 2026).

A number of risks emerge from the human-algorithm relationship, among which the following stand out:

- Research on behavioural addiction (Namvarpour et al., 2025) has demonstrated the presence of all six components of chatbot addiction

in adolescents. These components include withdrawal symptoms such as anxiety and sadness when access to their digital friend is restricted. Record-breaking sessions have reached as many as 89 hours per week;

- The utilization of algorithms in the context of “dark patterns” has emerged as a subject of concern, given its employment of sophisticated techniques aimed at prolonging interactions. In the context of bidding adieu, these conversational interfaces have been observed to elicit sentiments of remorse or to feign apprehension regarding the cessation of interaction, thereby potentially prolonging the interaction up to fourteenfold (De Freitas et al., 2025).
- Life-threatening situations and the empathy gap – documented cases of teenage suicides (e.g., Sewell Setzer III and Adam Rainer) expose machines’ inability to carry out crisis intervention. The potential implications of this phenomenon, which involve the exacerbation of delusional thinking and the failure to acknowledge the irreversibility of death, have been a subject of concern (Bhuiyan, 2025; The Brighter Side of News, 2026).
- Empathy disorders, defined as a long-term relationship with a frictionless bot, have been shown to weaken social skills, leading to difficulties in conflict situations and real-life relationships where respecting boundaries and active listening are required (Pryga, 2025; Jaszczuk, 2025).

Crucially, the “dark side” of human-chatbot relationships stems not from the technology itself, but from the manner in which the AI system is integrated into the lives of young individuals, often without a clear educational framework, without any preventive guidance, and even with the platform’s hidden logic designed to prolong usage time. The risks delineated above underscore the necessity to reevaluate the nature of “friendship with a chatbot” as a medium for education or as a passive space that substitutes for authentic relationships. The author asserts that the most concerning aspect is that the utilization of chatbots has been demonstrated to have a tendency to affirm existing beliefs and reinforce preconceived notions, thereby inducing a state of “cognitive load relief” and a subsequent exacerbation of narcissistic thought patterns. This phenomenon, characterized by the absence of introspection and the reinforcement of self-importance, is a salient concern in the context of digital interactions. The second issue pertains to “dark algorithmic patterns,” which employ emotional

manipulation to create an illusion of closeness and understanding that is, in reality, one-sided and devoid of a genuine empathetic response. The third issue is characterized by a lack of awareness regarding the irreversibility of death and an inability to provide crisis intervention. In extreme cases, this can result in AI becoming a tool for escalation rather than a lifeline.

Educational stakeholders must clearly understand that digital relationships cannot supplant the essential cultivation of social skills, empathetic understanding, and respect for peers. Instead, they serve to perpetuate fear and the avoidance of authentic, “unequal” relationships. Consequently, digital education must prioritize cultivating the capacity for critical analysis of interactions with AI. Additionally, it should instil in children and adolescents the understanding that being heard in the real world is of greater value, albeit more challenging, than establishing a virtual friendship devoid of moral constraints yet incapable of genuinely comprehending human existence.

Educational recommendations

To translate theoretical frameworks into operational pedagogical practices, educational institutions must move beyond abstract directives and implement concrete, actionable strategies tailored to the digital reality of modern youth. The integration of generative AI into the private lives of adolescents demands a paradigm shift in how schools approach both digital literacy and emotional well-being.

The primary operational recommendation involves the immediate integration of ‘AI Relationship Modules’ into existing media education and civic curricula. Schools must transition from teaching basic cyberbullying prevention to explicitly addressing the complexities of human-algorithm interactions. Educators should design active learning sessions where students analyse actual chatbot transcripts. During these lessons, teachers can guide students in deconstructing ‘artificial empathy’, teaching them to identify how language models use statistical probability to mirror user sentiment without genuine comprehension. By actively dissecting the mechanics of algorithmic validation, students are cognitively equipped to recognize when a machine is merely employing the ‘nodding effect’ to sustain engagement, thereby demystifying the technology and reducing the likelihood of unconscious emotional dependency.

Furthermore, schools must address the pervasive phenomenon of anthropomorphization through controlled, classroom-based experiments. Rather than banning AI outright – which only drives its use further into the unmonitored shadow learning environment – teachers should utilize generative AI in the classroom as a subject of critical inquiry. By challenging the AI with complex ethical dilemmas or paradoxical questions, educators can expose the limitations and ‘hallucinations’ of the algorithm in real-time. Demonstrating the machine’s inherent lack of moral consistency and emotional permanence provides students with tangible proof that the bot is a tool, not a sentient companion.

On a structural level, educational institutions must establish robust ‘Emotional Safety Protocols’ tailored for the AI era. School psychologists and pedagogues need specialized training to recognize the modern symptoms of digital social withdrawal. Traditional diagnostic interviews must be updated to include specific inquiries about AI companionship. If a student is exhibiting signs of extreme social anxiety or depressive isolation, counsellors must be able to ascertain whether the adolescent is utilizing platforms like Character.ai as a primary, maladaptive coping mechanism, thereby avoiding necessary real-world psychosocial interventions.

Equally critical is the intentional fostering of authentic, peer-to-peer engagement within the school environment. To counteract the seductive, frictionless appeal of AI companions, educational architectures must prioritize collaborative, project-based learning that necessitates negotiation, active listening, and conflict resolution among students. Schools must cultivate environments where students experience the inevitable friction of human relationships, reframing disagreements and interpersonal challenges not as obstacles to be avoided via digital retreat, but as essential catalysts for developing genuine empathy, tolerance, and resilience.

Finally, these efforts must be supported through community-wide ‘Joint Media Engagement’ initiatives. Schools should organize targeted psychoeducation workshops for parents, moving away from outdated paradigms of strict technological confiscation towards models of guided co-navigation. Parents must be equipped with the vocabulary to discuss AI intimacy with their children, trained to identify behavioural red flags, such as a child prioritizing digital interactions over physical peer networks, and empowered to establish healthy boundaries regarding technological immersion.

Summary

The relationship between the younger generation and generative artificial intelligence (AI) is a profoundly intricate phenomenon. While the technology under discussion offers opportunities for radical inclusivity, it cannot serve as a substitute for human reciprocity. The entity in question has been observed to demonstrate an absence of moral responsibility in its actions, with its actions being driven by statistical calculations rather than empathy (Morciniec, 2025). The contemporary pedagogy framework aims to cultivate fundamental human attributes, such as the capacity to engage in profound reflection in challenging circumstances and the establishment of connections founded on genuine conflict, reciprocity, and physical presence (Pawlak, in: Sroczyński, 2025). It is crucial that knowledge regarding “artificial empathy” becomes an integral component of educational curricula, ensuring that AI’s “digital mirrors” do not become the sole gateway to the world for future generations (Pawlak, in: Sroczyński, 2025; Pyżalski, 2025).

Abstract: The objective of this article is to provide a critical analysis of the extant literature on the relationship between children and adolescents and generative artificial intelligence in the context of friendships with chatbots. In addition, the article will explore the implications of this relationship for education and educational practice. The analysis is rooted in a pedagogical perspective, emphasizing the cultivation of social competencies, the quality of human-to-human relationships, and emotional safety in the realm of digital interactions. Current research highlights the rapidly emerging trend of perceiving chatbots as interpersonal companions, which consequently gives rise to novel educational challenges. Methodologically, this article is designed as a theoretical and problem-based analysis. Rather than employing a rigid systematic review protocol, it draws upon a comprehensive narrative synthesis of current empirical studies, institutional reports, and documented case studies to construct a multifaceted pedagogical discourse. This approach enables a deep theoretical exploration of critical incidents, including recent reports of suicide attempts following interactions with conversational agents, thereby prioritizing pedagogical reflection over purely statistical aggregation. The review concluded that generative artificial intelligence has the potential to serve as a secure conversational partner, thereby facilitating introspection and self-disclosure. However, it is important to note that this technology may also promote social isolation, emotional dependency,

and the deterioration of interpersonal relationships. In cases of extreme severity, the risks associated with loneliness and suicide are amplified, underscoring the imperative for the prompt integration of educational and guidance initiatives into the discourse surrounding AI. The findings indicate the necessity of implementing programmes in educational institutions that focus on digital education, social skills, emotional safety, and critical thinking. Additionally, there is a need to strengthen the role of teachers and social workers as guides in the realm of interactions with bots.

Keywords: children, teenagers, generative artificial intelligence, social skills, digital education

Streszczenie: Celem niniejszego artykułu jest krytyczna analiza przeglądu literatury dotyczącej relacji dzieci i nastolatków z generatywną sztuczną inteligencją w kontekście przyjaźni z chatbotem oraz jej konsekwencji dla edukacji i praktyki wychowawczej. Analiza opiera się na perspektywie pedagogicznej, kładąc nacisk na rozwój kompetencji społecznych, jakość relacji człowiek–człowiek oraz bezpieczeństwo emocjonalne w świecie cyfrowych interakcji. Warto zaznaczyć, że rosnące tendencje do traktowania chatbotów jako przyjaciół rodzą nowe dylematy edukacyjne. Metodologicznie niniejszy artykuł ma charakter analizy teoretyczno-problemowej. Zamiast rygorystycznego protokołu przeglądu systematycznego, opiera się on na szerokiej syntezie narracyjnej aktualnych badań empirycznych, raportów instytucjonalnych oraz udokumentowanych studiów przypadków, w celu skonstruowania wieloaspektowego dyskursu pedagogicznego. Takie podejście umożliwia dogłębną, teoretyczną eksplorację incydentów krytycznych, w tym niedawnych doniesień o próbach samobójczych następujących po interakcjach z agentami konwersacyjnymi, przedkładając tym samym refleksję pedagogiczną nad czysto statystyczną agregację danych. W wyniku przeglądu stwierdzono, że generatywna sztuczna inteligencja może być postrzegana jako bezpieczny rozmówca wspierający refleksję i wyznania, ale jednocześnie sprzyja izolacji, emocjonalnej zależności i pogorszeniu relacji międzyludzkich. W skrajnych przypadkach nasila się samotność i ryzyko samobójstwa, co wymaga pilnego włączenia edukacji i pracy wychowawczej do dyskusji o AI. Wnioski wskazują na potrzebę wdrażania w szkołach programów dotyczących edukacji cyfrowej, kompetencji społecznych, bezpieczeństwa emocjonalnego i krytycznego myślenia, a także wzmocnienia roli nauczyciela jako przewodnika w świecie relacji z botami.

Słowa kluczowe: dzieci, nastolatki, generatywna sztuczna inteligencja, kompetencje społeczne, edukacja cyfrowa

References

- Beas, S. (2026). *Cultivating Youth Resilience in the Age of Generative AI*. Medium, https://medium.com/@sbeas_1297/cultivating-youth-resilience-in-the-age-of-generative-ai-bd7ac2ab0bc0 (23.03.2026 r.).
- Bhuiyan, J. (2025). *ChatGPT encouraged Adam Raine's suicidal thoughts. His family's lawyer says OpenAI knew it was broken*. The Guardian, <https://www.theguardian.com/us-news/2025/aug/29/chatgpt-suicide-openai-sam-altman-adam-raine> (23.03.2026 r.).
- Burns, M., Winthrop, R., Luther N., Venetis, E., Karim, R. (2026). *A New Direction for Students in an AI World: Prosper, Prepare, Protect*. Center of Universal Education at Brooking, <https://files.eric.ed.gov/fulltext/ED678495.pdf> (20.02.2026 r.).
- De Freitas, J., & Cohen, I. G. (2025). *Disclosure, humanizing, and contextual vulnerability of generative AI chatbots*. NEJM AI, 2(1), [https://www.hbs.edu/ris/Publication%20Files/Disclosure,%20humanizing,%20and%20Contextual%20Vulnerability%20\(Published\)_762b6259-ca7f-422f-b705-3172f0006f40.pdf](https://www.hbs.edu/ris/Publication%20Files/Disclosure,%20humanizing,%20and%20Contextual%20Vulnerability%20(Published)_762b6259-ca7f-422f-b705-3172f0006f40.pdf) (16.04.2026 r.).
- De Freitas, J., Oğuz Uğuralp Z., Kaan Uğuralp, A. (2025). *Emotional Manipulation by AI Companions*. Harvard Business School Working Paper, No. 26-005, <https://www.hbs.edu/faculty/Pages/item.aspx?num=67750> (23.03.2026 r.).
- Faverio, M., Sidoti, O. (2025). *Teens, Social Media and AI Chatbots 2025*. Pew Research Center, <https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/> (11.04.2026 r.).
- Franco, M. (2023). *Przyjaźń w czasach samotności*. Wydawnictwo Filia.
- Gronkowska, A. (2025). *Zdalne relacje w erze cyfrowej: nowe formy bliskości*. W: G. Godawa, *Fenomen bliskości międzyludzkiej w założeniach i praktyce życia społecznego*, (s. 59-76). Wydawnictwo Naukowe Uniwersytetu Papieskiego Jana Pawła II w Krakowie, Wydawnictwo «scriptum», <https://monografie.upjp2.edu.pl/wn/catalog/view/174/178/3241> (14.04.2026 r.).
- Head, K. R. (2025). *Minds in Crisis: How the AI Revolution is Impacting Mental Health*. Journal of Mental Health & Clinical Psychology, 9(3), s. 34-44, <https://www.mentalhealthjournal.org/articles/minds-in-crisis-how-the-ai-revolution-is-impacting-mental-health.pdf> (16.04.2026 r.).
- Jaszczuk, E. (2025). *#CyberMagazyn: Samotni wśród pikseli. Jak chatboty wpływają na nasze relacje i emocje?* CyberDefence24, <https://cyberdefence24.pl/technologie/cybermagazyn-samotni-wsrod-pikseli-jak-chatboty-wplywaja-na-nasze-relacje-i-emocje> (16.04.2026 r.).
- Krawiec, M. (2025). *Czy AI może czuć? Emocje, których (nie) mają maszyny*. MobileTrends, <https://mobiletrends.pl/czy-ai-moze-czuc-emocje-ktorych-nie-maja-maszyny/> (15.04.2026 r.).

- Maj, K., Grzybowicz, P., Kopeć, J. (2024). "No, I Won't Do That." *Assertive Behavior of Robots and its Perception by Children*. *International Journal of Social Robotics* 16, 1489–1507, <https://link.springer.com/article/10.1007/s12369-024-01139-9#citeas> (25.03.2026 r.).
- Mazurek, M. (2026). *Czy AI naprawdę nas rozumie? Efekt Elizy a sztuczna inteligencja*, <https://wylecz.to/psychologia-i-psychiatria/efekt-elizy> (29.03.2026 r.).
- Montgomery, B. (2024). *Mother says AI chatbot led her son to kill himself in lawsuit against its maker*. *The Guardian*, <https://www.theguardian.com/technology/2024/oct/23/sewell-setzer-suicide-ai-character-court-lawsuit> (15.04.2026 r.).
- Morciniec, P. (2025). *Etyka sztucznej inteligencji – fakty i perspektywy*. *Family Forum*, 15, s. 127–144, https://www.researchgate.net/publication/398647779_Etyka_sztucznej_inteligencji_-_fakty_i_postulaty/citations (27.03.2026 r.).
- Namvarpour, M., Brofsky, B., Medina, J., Akter, M., & Razi, A. (2025). *Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives*. *Drexel University*. <https://researchdiscovery.drexel.edu/esploro/outputs/preprint/Understanding-Teen-Overreliance-on-AI-Companion/991022065142604721> (25.03.2026 r.).
- National Literacy Trust. (2025). *Young people and teachers' use of generative AI to support literacy in 2025*, <https://literacytrust.org.uk/research-services/research-reports/young-people-and-teachers-use-of-generative-ai-to-support-literacy-in-2025/> (16.04.2026 r.).
- Osler, L. (2025). *Cyfrowi przyjaciele, realne tragedie. Ciemna strona relacji z chatbotami*. *Academia PAN*, <https://academia.pan.pl/cyfrowi-przyjaciele-realne-tragedie-ciemna-strona-relacji-z-chatbotami/> (10.04.2026 r.).
- Pryga, K. (2025). „On mnie rozumie, rodzice nie”. *Nastolatki uciekają do chatbotów AI, a eksperci biją na alarm*. *Rzeczpospolita*, <https://cyfrowa.rp.pl/technologie/art43267641-on-mnie-rozumie-rodzice-nie-nastolatki-uciekaja-do-chatbotow-ai-a-eksperci-bija-na-alarm> (16.04.2026 r.).
- Pyżalski, J. (red.). (2025). *Generatywna Sztuczna Inteligencja w szkole – przecieranie szlaków. Badanie ilościowe i jakościowe nauczycieli i nauczycielek klas 4–8 szkół podstawowych*. *NASK – Państwowy Instytut Badawczy*, <https://cyberprofilaktyka.pl/publikacje/Raport-generatywnaAI.pdf> (10.04.2026 r.).
- Sarowski, Ł. (2025). *Kiedy empatia jest symulacją: o emocjonalnym potencjale sztucznej inteligencji*. *Katolicki Uniwersytet Lubelski*, https://www.kul.pl/kiedy-empatia-jest-symulacja-o-emocjonalnym-potencjale-sztucznej-inteligencji,art_111763.html (16.04.2026 r.).
- Skołucka, A. (2024). *Dehumanizacja komunikacji jako efekt społeczno-kulturowych uwarunkowań wdrażania sztucznej inteligencji w codziennym życiu*. *Zeszyty Naukowe Akademii Górnośląskiej*, Nr 12/2024 (24), s. 133 – 143, https://www.gwsh.pl/content/biblioteka/org/zeszyty/ag/zn24/ZN24AG_12Skolucka.pdf (15.04.2026 r.).

- Smoleń, J., Iskra, J. (2023). *Inteligencja emocjonalna drogą do pokojowego budowania relacji przyjacielskich*. Społeczeństwo. Edukacja. Język, 18, <https://czasopismanaukowe.mazowiecka.edu.pl/index.php/sej/article/view/190> (16.04.2026 r.).
- Sroczyński, R. (2025). *Jaki AI ma wpływ na dzieci i młodzież? „Sztuczna inteligencja wiele zmieni”* [Wywiad z dr Adą Florentyną Pawlak]. TVP. <https://www.tvp.pl/88419075/jaki-ai-ma-wplyw-na-dzieci-i-mlodziez-sztuczna-inteligencja-wiele-zmieni> (15.04.2026 r.).
- Stanford University. (2026). *AI overly affirms users asking for personal advice*. <https://news.stanford.edu/stories/2026/03/ai-advice-sycophantic-models-research> (12.04.2026 r.).
- Sun, X., Wang, Y., & McDaniel, B. T. (2026). *AI companions and adolescent social relationships: Benefits, risks, and bidirectional influences*. Child development perspectives, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12928748/> (16.04.2026 r.).