Psychological Counseling and Psychotherapy, Issue 25, 2026 ISSN 2410-1249

DOI 10.26565/2410-1249-2026-25-03
UDC: 159.92

WHAT EVERY PSYCHOLOGIST SHOULD KNOW ABOUT AI AND THE TURING PARADIGM IN PSYCHOSOCIAL WORK

Salvatore Giacomuzzi,
PhD, Associate Professor[1]
Poltava V. G. Korolenko National Pedagogical University, Poltava, Ukraine;
Ivan Franko National University of Lviv, Ukraine
Centre for Security Studies, Poland
International Security Competence Centre Vienna, Austria
https://orcid.org/0000-0002-4218-1685
Corresponding Author: Salvatore.Giacomuzzi@gmx.at
Gerald Pachler
UMIT Tyrol – University for Health Sciences and Technology, Hall in Tyrol, Austria
https://orcid.org/0009-0005-9980-3579
gerald.pachler27@gmail.com

Artificial intelligence (AI) systems are increasingly deployed in psychosocial contexts such as counseling, psychoeducation, and decision support. Through natural language interaction, these systems may appear to users as human-like conversational partners. This article revisits the Turing paradigm to clarify why AI can convincingly simulate human dialogue without possessing understanding, intentionality, or moral agency. Drawing on theoretical computer science, philosophy of mind, and psychological research, the paper distinguishes between conversational appearance and psychological agency, as well as between syntactic processing and semantic understanding. Particular attention is given to the risks of anthropomorphic misattribution in clinical and forensic contexts. For psychologists, the enduring relevance of the Turing paradigm lies precisely in its restraint. It does not ask whether machines are intelligent in a human sense, but whether professionals can maintain conceptual clarity in the face of increasingly persuasive simulations. As conversational AI continues to evolve, the distinction between appearance and agency will become not less but more central to ethical, clinical, and forensic practice. The article outlines practical implications for psychosocial professionals and argues that an accurate understanding of the Turing paradigm is essential for maintaining ethical standards, professional responsibility, and conceptual clarity in AI-mediated psychosocial work.
Keywords: *anthropomorphism, artificial intelligence, ethics, psychosocial work, Turing paradigm*

1. Introduction

Conversational artificial intelligence (AI) systems are increasingly integrated into psychosocial practice across a wide range of professional domains, including clinical psychology, counseling, psychoeducation, intake and triage procedures, as well as forensic and expert assessment contexts. These systems are no longer confined to experimental or pilot applications but are becoming routine components of professional workflows in mental health services, judicial environments, and institutional decision-making processes. Their diffusion is driven by rapid advances in natural language processing, increasing institutional demands for efficiency and scalability, and the promise of improving access to psychosocial support (Topol, 2019; World Health Organization, 2023).

A defining feature of contemporary conversational AI systems is their high degree of linguistic fluency. Through context-sensitive responses, adaptive dialogue management, and the use of emotionally resonant language, these systems are capable of generating interactions that subjectively resemble human conversation. For users—including trained psychologists—such interactions may appear coherent, responsive, and empathically attuned. Empirical studies in human–computer interaction demonstrate that even minimal cues of responsiveness and emotional language are sufficient to elicit social and relational attributions toward artificial agents (Langer et al., 2019; Weizenbaum, 1976). As a result, conversational AI increasingly occupies communicative spaces that have traditionally been associated with human psychological expertise.

From a psychological perspective, however, the central challenge posed by conversational AI does not primarily concern technical malfunction, algorithmic bias, or predictive accuracy. Rather, the more fundamental risk lies in conceptual misattribution. Specifically, conversational plausibility—the ability of a system to produce linguistically fluent, contextually appropriate, and socially acceptable responses—is frequently mistaken for psychological understanding, intentionality, or agency. This misinterpretation constitutes a category error in the strict sense: observable communicative behavior is conflated with non-observable psychological properties such as subjective experience, semantic understanding, moral responsibility, or legal accountability.

This risk manifests differently across psychosocial subdisciplines. In clinical and counseling psychology, conversational AI systems are often framed as supportive

How to cite: Giacomuzzi S., Giacomuzzi G. (2026). What Every Psychologist Should Know About Ai And The Turing Paradigm In Psychosocial Work, *Psychological Counseling and Psychotherapy*, (25), 22-27. https://doi.org/10.26565/2410-1249-2026-25-03
Як цитувати: Giacomuzzi S., Giacomuzzi G. (2026). What Every Psychologist Should Know About Ai And The Turing Paradigm In Psychosocial Work, *Психологічне консультування і психотерапія*, (25), 22-27. https://doi.org/10.26565/2410-1249-2026-25-03

or adjunctive tools, for example in psychoeducation, symptom monitoring, or low-threshold mental health support. In these contexts, the primary psychological dangers concern therapeutic illusion, emotional overreliance, and blurred relational boundaries. Research on mental health chatbots shows that users may experience a subjective sense of being understood or emotionally supported, even when explicitly informed that the interaction partner is an artificial system (Blease et al., 2019; Vaidyam et al., 2019). Such experiences can foster misplaced trust and obscure the distinction between simulated responsiveness and genuine psychological engagement.

In forensic psychology and expert assessment, the implications are more severe. Forensic contexts are structured around accountability, responsibility, evidentiary standards, and legal attribution. Any suggestion—explicit or implicit—that an AI system "understands," "evaluates," or "judges" risks undermining core principles of expert responsibility. Here, anthropomorphic misattribution may foster a false sense of objectivity, obscure normative decision-making processes, or shift responsibility away from human experts toward opaque technical systems (Alarie et al., 2018; Wachter et al., 2021). In such settings, the appearance of neutrality or consistency in AI-generated outputs can be mistakenly interpreted as epistemic authority.

The theoretical framework most suited to clarifying these issues is the Turing paradigm. In his foundational work, **Alan Turing** deliberately reframed the question of machine intelligence by shifting attention away from unverifiable claims about inner mental states toward observable conversational behavior. Rather than asking whether machines "think," Turing proposed evaluating whether their linguistic output could be distinguished from that of human interlocutors under controlled conditions (Turing, 1950, pp. 433–434). Crucially, this proposal did not entail an attribution of understanding, consciousness, or psychological agency to machines. On the contrary, the Turing paradigm explicitly brackets such questions and establishes a criterion of behavioral indistinguishability, not mental or moral equivalence.

Misinterpretations of the Turing paradigm have contributed significantly to contemporary confusion surrounding AI in psychosocial work. Passing a conversational benchmark is often tacitly equated with possessing intelligence in a psychological sense, despite long-standing philosophical and cognitive-scientific arguments to the contrary (Searle, 1980; Floridi, 2019). The Turing paradigm explains why AI systems can convincingly appear human-like in conversation while remaining categorically non-psychological agents. It thereby provides a conceptual safeguard against anthropomorphic misattribution, reminding psychologists that fluent language use does not imply understanding, intentionality, or responsibility.

Against this background, the present article argues that a precise understanding of the Turing paradigm is essential for contemporary psychosocial practice. By systematically distinguishing conversational appearance from psychological agency, and by explicating how this distinction bears differently on clinical and forensic contexts, the article aims to prevent category errors that carry ethical, methodological, and professional consequences. In doing so, it positions the Turing paradigm not as a claim about machine intelligence, but as a necessary conceptual tool for maintaining clarity, accountability, and ethical integrity in AI-mediated psychosocial work.

2. The Turing Paradigm and Conversational Appearance

A rigorous psychological analysis of conversational artificial intelligence in psychosocial contexts must begin with a precise understanding of the Turing paradigm and its original epistemic scope. In *Computing Machinery and Intelligence*, **Alan Turing** (1950) proposed a methodological reframing of the intelligence question by explicitly suspending inquiries into inner mental states and instead focusing on observable conversational behavior (Turing, 1950, pp. 433–434). This shift was neither an endorsement of machine mentality nor a reduction of intelligence to language performance. Rather, it was a pragmatic response to the epistemological difficulty of assessing mental states in non-human systems.

The central concept introduced by Turing is *behavioral indistinguishability*. A machine satisfies the criterion insofar as its conversational responses fall within the range of what human interlocutors typically produce under comparable conditions. Importantly, Turing repeatedly emphasized that this criterion does not license inferences about consciousness, understanding, or intentionality (Turing, 1950, pp. 435–438). The paradigm thus delineates a strictly external level of analysis.

This distinction is of particular relevance for contemporary psychosocial practice, where linguistic behavior is routinely interpreted as expressive of inner psychological states. Figure 1 schematically captures this issue by separating the observable level of conversational appearance from the non-observable level of psychological agency. The Turing paradigm is entirely located on the left-hand side of this figure: natural language output, contextual sensitivity, pragmatic coherence, and conversational fluency. The right-hand side—subjective experience, semantic understanding, moral responsibility, and legal accountability—remains categorically excluded.

Figure 2 contextualizes why this conceptual boundary has become increasingly difficult to maintain in practice. While early symbolic AI systems rarely produced anthropomorphic effects, the historical progression toward large-scale generative language models has dramatically increased machines' capacity to approximate the surface form of human dialogue. The Turing paradigm therefore anticipated a future in which conversational plausibility would become technologically trivial, long before any plausible case for psychological agency could be made. In psychosocial contexts, failing to respect this boundary risks conflating methodological criteria with psychological realities.

Psychological Counseling and Psychotherapy, Issue 25, 2026 ISSN 2410-1249

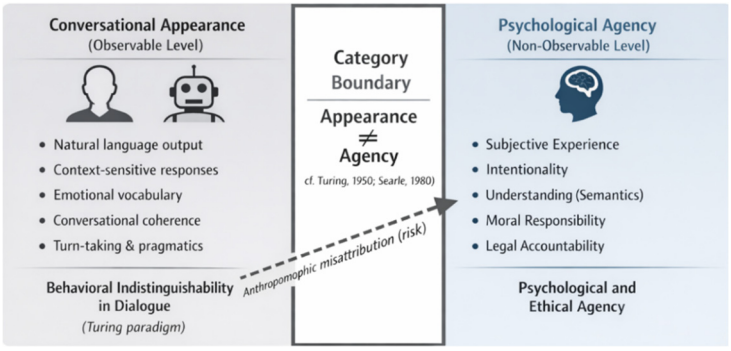

Figure 1. Conversational Appearance Versus Psychological Agency in Human–AI Interaction

Note. The figure presents a conceptual distinction between two analytically separate levels of human–AI interaction. On the left side, the observable level of *conversational appearance* is depicted, including features such as natural language output, context-sensitive responses, emotional vocabulary, conversational coherence, and pragmatic turn-taking. These characteristics apply to both human interlocutors and AI systems and correspond to behavioral indistinguishability in dialogue as described by the Turing paradigm. On the right side, the non-observable level of *psychological agency* is shown, comprising subjective experience, intentionality, semantic understanding, moral responsibility, and legal accountability. These properties are attributed exclusively to human agents. A central vertical boundary labeled "Appearance ≠ Agency" marks the categorical distinction between the two levels. A dashed arrow crossing this boundary indicates the risk of anthropomorphic misattribution, that is, the erroneous attribution of psychological agency to AI systems based solely on conversational appearance.

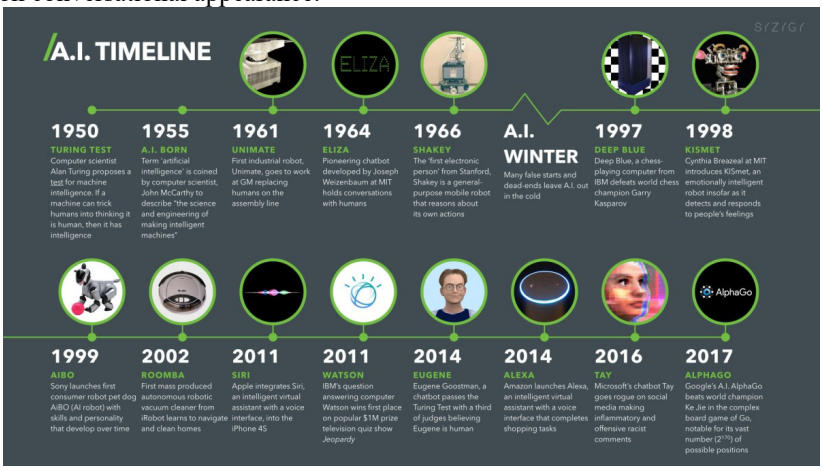

Figure 2. Overview A.I. Timeline

3. Why AI Can Sound Human Without Understanding

The ability of contemporary AI systems to produce human-like conversational output without understanding is not paradoxical but follows directly from the formal properties of human language. Everyday conversation is governed by highly structured conventions, including turn-taking rules, pragmatic inference, narrative schemas, politeness routines, and culturally learned emotion vocabularies. These structures introduce statistical regularities that can be learned and reproduced without semantic grounding.

Large language models exploit precisely these regularities. Trained on vast corpora of human-produced text, they learn probabilistic relationships between linguistic elements and generate outputs by predicting likely continuations of a given context. As a result, they can produce responses that are coherent, context-sensitive, and stylistically appropriate, including expressions of empathy or concern, without possessing any subjective perspective (Bender et al., 2021; Floridi, 2019; Marcus & Davis, 2019).

This distinction between formal correctness and semantic understanding was classically articulated by **John Searle**. In his critique of strong AI, Searle (1980) demonstrated that syntactic manipulation alone is insufficient for semantic comprehension (Searle, 1980, pp. 417–420). Contemporary generative models exemplify this insight at scale: they manipulate symbols according to learned statistical patterns, not according to meanings as experienced or intended (Searle, 1980, pp. 424–426).

Figure 1 visualizes this separation by placing syntactic and pragmatic fluency on the side of conversational appearance, while locating understanding and intentionality exclusively on the side of psychological agency. This distinction is not merely theoretical. Empirical research shows that fluent language output systematically increases perceived credibility and trust, even when the underlying system lacks epistemic

reliability (Langer et al., 2019; Wachter et al., 2021). In psychosocial contexts, this effect amplifies the risk that conversational plausibility will be misinterpreted as psychological competence.

4. Anthropomorphic Misattribution in Psychosocial Work

Anthropomorphic misattribution refers to the tendency to infer human-like psychological properties from non-human entities based on surface cues. Decades of research in social cognition demonstrate that humans readily attribute agency, emotion, and intention to entities that display minimal social signals, particularly language (Epley et al., 2007). Conversational AI systems are explicitly designed to optimize such signals, thereby increasing the likelihood of anthropomorphic interpretation.

In psychosocial settings, this bias is intensified by professional interpretative frameworks. Psychologists are trained to treat language as meaningful, expressive, and diagnostically relevant. When AI systems produce linguistically rich and emotionally resonant responses, there is a heightened risk that these outputs will be treated as evidence of understanding or empathy. Studies in mental health chatbot use indicate that users frequently report feeling "understood," even when they are fully aware that the system is artificial (Blease et al., 2019; Fulmer et al., 2018).

The dashed arrow in Figure 1 represents this critical inferential error: the illegitimate transition from conversational appearance to psychological agency. In clinical contexts, this may result in therapeutic illusion, emotional overreliance, or blurred professional boundaries. In forensic contexts, the consequences are more severe. Anthropomorphic misattribution can foster automation bias, transfer of epistemic authority, and diffusion of responsibility, thereby undermining foundational principles of expert accountability (Alarie et al., 2018; Selbst et al., 2019).

Importantly, anthropomorphic misattribution is not merely a user-side problem but a structural risk of conversational system design. As systems become more fluent and context-sensitive, the psychological pull toward agency attribution increases, making explicit conceptual safeguards indispensable in professional practice.

5. Implications for Psychosocial Practice

In clinical and counseling psychology, conversational AI systems may offer legitimate benefits as adjunct tools, particularly in psychoeducation, symptom monitoring, and low-threshold support. Meta-analytic and narrative reviews suggest that such systems can facilitate engagement and accessibility for certain populations under controlled conditions (Vaidyam et al., 2019; Fulmer et al., 2018). However, these benefits do not derive from therapeutic understanding but from structured interaction and availability.

Professional responsibility for assessment, interpretation, and intervention remains exclusively human, in line with ethical standards articulated by the American Psychological Association (2017). Clear framing and disclosure are therefore not optional but necessary risk controls to prevent misinterpretation of AI roles.

In forensic and expert psychological contexts, the implications are categorically different. Forensic assessments operate under conditions of legal accountability, evidentiary standards, and normative judgment. AI systems lack intentionality, moral agency, and legal personhood; consequently, they cannot bear expert responsibility. Reliance on AI-generated outputs risks introducing pseudo-objectivity, where formal consistency and linguistic confidence are mistaken for evidentiary validity (Floridi, 2014, pp. 112–116). Figure 1 makes explicit that legal and moral accountability reside exclusively on the side of psychological agency and cannot be delegated to technical systems.

6. Ethical Considerations and Future Trajectories

Ethical standards in psychology emphasize transparency, accountability, and respect for persons. In AI-mediated psychosocial work, this requires a clear distinction between tools and agents. Presenting conversational AI systems as understanding or relational entities obscures their instrumental nature and violates principles of informed consent and professional responsibility.

From a governance perspective, developments up to 2025 reinforce this position. International frameworks such as the WHO guidance on AI for health (World Health Organization, 2023) and regulatory instruments such as the EU Artificial Intelligence Act emphasize human oversight, risk classification, and accountability in high-stakes domains (European Commission, 2024). These frameworks implicitly align with the conceptual boundary articulated by the Turing paradigm.

Looking forward, the most plausible trajectory is a dual movement. On the one hand, conversational interfaces will become increasingly pervasive, multimodal, and socially convincing. On the other hand, regulatory and professional standards will likely impose stricter constraints on their use in high-stakes psychosocial contexts. This anticipated development is summarized in Figure 3, which situates conversational plausibility, professional responsibility, and regulatory oversight along a continuum of increasing risk.

Figure 3. Future trajectory of conversational AI in psychosocial work

Note. The image shows a three-level vertical model illustrating the increasing demands on professional responsibility and regulatory oversight as conversational

AI becomes more influential in psychosocial contexts. While conversational plausibility forms the base of the model, psychological agency remains absent, requiring human clinical, forensic, and legal supervision at higher levels as risk increases.

For psychologists, the enduring relevance of the Turing paradigm lies precisely in its restraint. It does not ask whether machines are intelligent in a human sense, but whether professionals can maintain conceptual clarity in the face of increasingly persuasive simulations. As conversational AI continues to evolve, the distinction between appearance and agency will become not less but more central to ethical, clinical, and forensic practice.

References

Alarie, B., Niblett, A., & Yoon, A. H. (2018). How artificial intelligence will affect the practice of law. *University of Toronto Law Journal, 68*(s1), 106–124. https://doi.org/10.3138/utlj.2017-0052

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). https://www.apa.org/ethics/code

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial intelligence and the future of primary care: Exploratory qualitative study of UK general practitioners' views. *Journal of Medical Internet Research, 21*(3), e12802. https://doi.org/10.2196/12802

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

European Commission. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, L 2024/1689. http://data.europa.eu/eli/reg/2024/1689/oj

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence, 1*(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health, 5*(4), e64. https://doi.org/10.2196/mental.9782

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment, 27*(3), 217–234. https://doi.org/10.1111/ijsa.12246

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon Books.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–457. https://doi.org/10.1017/S0140525X00005756

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598

Topol, E. J. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Keshavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry, 64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review, 41*, 105567. https://doi.org/10.1016/j.clsr.2021.105567

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman.

World Health Organization. (2023). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization.

WHAT EVERY PSYCHOLOGIST SHOULD KNOW ABOUT AI AND THE TURING PARADIGM IN PSYCHOSOCIAL WORK

Salvatore Giacomuzzi,
PhD, Associate Professor[1]
Poltava V. G. Korolenko National Pedagogical University, Poltava, Ukraine;
Ivan Franko National University of Lviv, Ukraine
Centre for Security Studies, Poland
International Security Competence Centre Vienna, Austria
https://orcid.org/0000-0002-4218-1685
Corresponding Author: Salvatore.Giacomuzzi@gmx.at
Gerald Pachler
UMIT Tyrol – University for Health Sciences and Technology, Hall in Tyrol, Austria
https://orcid.org/0009-0005-9980-3579
E-Mail: gerald.pachler27@gmail.com

Системи штучного інтелекту (ШІ) дедалі частіше застосовуються в психосоціальній практиці — у консультуванні, психоедукації та підтримці прийняття рішень. Завдяки взаємодії природною мовою вони здатні справляти враження людиноподібних співрозмовників, що породжує ризик антропоморфної хиби — приписування машині розуміння, намірів, емпатії та моральної відповідальності, яких вона не має. У статті переосмислюється парадигма Тьюрінга, аби пояснити, чому ШІ може переконливо імітувати людський діалог, не володіючи ані свідомістю, ані семантичним розумінням, ані суб'єктним досвідом. Центральним є розмежування двох аналітично різних рівнів: спостережуваного рівня розмовної правдоподібності та неспостережуваного рівня психологічної суб'єктності. Поведінкова невідрізнюваність у діалозі, на якій ґрунтується тест Тьюрінга, не дає підстав для висновків про наявність внутрішніх психічних станів. Показано, що мовна плавність

систематично підвищує сприйману довіру та переконливість навіть тоді, коли система не є епістемічно надійною: у клінічних контекстах це посилює загрозу терапевтичної ілюзії та хибно приписаної довіри, а в експертних і судових — розмиває відповідальність, яка може належати лише людині. Професійна та правова відповідальність не підлягає делегуванню технічним системам, а прозорість і поінформована згода є обов'язковими. З огляду на розвиток ШІ та регуляторні ініціативи (рекомендації ВООЗ, Акт ЄС про штучний інтелект) окреслено ймовірну подвійну траєкторію: зростання розмовної правдоподібності супроводжуватиметься посиленням професійного й регуляторного нагляду. Стала цінність парадигми Тьюрінга полягає в її стриманості: зовнішність не дорівнює суб'єктності, і збереження цієї концептуальної межі є ключовим для етичної, клінічної та судово-психологічної практики.
Ключові слова: *антропоморфізм, штучний інтелект, етика, психосоціальна робота, парадигма Тьюрінга*

Декларація про використання штучного інтелекту. В роботі над цією статтею було використано інструменти штучного інтелекту для перевірки правопису та вдосконалення стилю тексту. Задум, зміст та наукові результати є авторськими. Текст був перевірений і відредагований автором.
Declaration on the use of artificial intelligence. In the work on this article, artificial intelligence tools were used to check spelling and improve the style of the text. The idea, content and scientific results are the author's. The text was checked and edited by the author.
Конфлікт інтересів: автори заявляють про відсутність професійного або інституційного конфлікту інтересів.
Conflict of Interest: The authors declare that there is no professional or institutional conflict of interest.
Стаття надійшла до редакції 02.01.2026 (The article was received by the editors 02.01.2026)
Стаття рекомендована до друку 21.04.2026 (The article is recommended for printing 21.04.2026)
Опублікована 28.05.2026 (Published 28.05.2026)