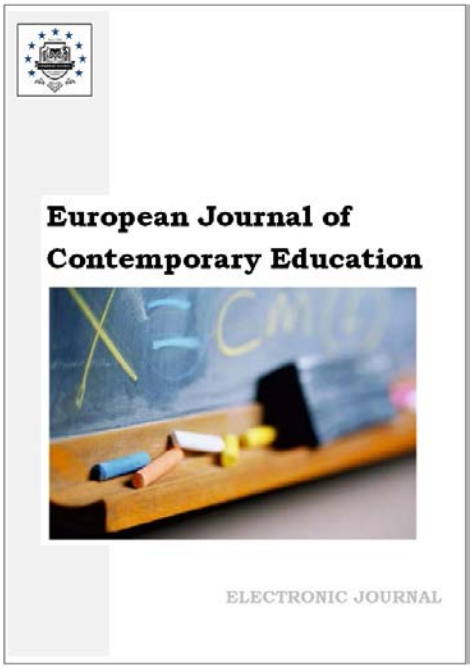
European Journal of Contemporary Education
E-ISSN 2305-6746
2026. 15(1): 76-89
DOI: 10.13187/ejced.2026.1.76
https://ejce.cherkasgu.press



A Comparative Analysis of Interpretation Strategies between Psychology Students and Artificial Intelligence in an Educational Context

Tetyana Ivanova [a, *], Andrii E. Lebid [a, b], Yuliia Typtiuk [a]

[a] Sumy State University, Sumy, Ukraine
[b] Cherkas Global University, Houston, USA

Abstract

The aim of this study is to explore the potential and limitations of artificial intelligence (AI) in psychological counseling. A comparative method was used, based on assessing the differences in interpretation between psychology students and AI systems. Specifically, the phenomenon of overinterpretation, in which conclusions lack sufficient empirical support in the source material, was analyzed.

This article utilized a mixed-methods approach. Specifically, the theoretical part included a literature review on AI counseling across five parameters: methodological commitment, emotional support, therapeutic alliance, ethical considerations, and accessibility. The empirical part involved a pilot pedagogical experiment in which psychology students (N = 44) and the Claude 3.5 Sonnet system independently analyzed identical psychological interviews (N = 22). The resulting analytical texts (N = 66) were subjected to discourse analysis based on the identified markers. For a more in-depth analysis, an independent expert review was used. A theoretical review found that AI is effective within structured protocols (e.g., cognitive behavioral therapy). However, AI has several limitations in establishing a therapeutic alliance and interpreting deep emotional experiences. An empirical study identified stylistic differences between student-produced texts and AI-generated texts. Specifically, student-produced texts were more than three times more likely to contain markers of epistemic caution (3.84 versus 1.20 per 1,000 words). AI-generated texts showed a twofold increase in markers of causality (4.58 versus 1.92) and recommendations (3.61 versus 1.60). Furthermore, AI was more likely to overinterpret type A texts, drawing categorical conclusions based on the absence of information in the original text. Thus, integrating AI into psychology education requires the development of a new professional competency, the essence of which lies in the ability to critically evaluate AI-generated content. Also important is the ability to identify interpretive errors and apply strict evidence boundaries. Artificial intelligence

* Corresponding author
E-mail addresses: t.ivanova@socio.sumdu.edu.ua (T. Ivanova)

can serve as an effective didactic tool for developing critical thinking, provided it is used as a supplementary educational resource. Using artificial intelligence as a standard for professional analysis is not pedagogically incorrect.

Keywords: artificial intelligence, psychological counseling, AI counseling, professional psychology education, overinterpretation, critical thinking, AI literacy.

1. Introduction

The capabilities of artificial intelligence are currently expanding. Artificial intelligence (AI) enables the performance of tasks that have traditionally required human intellectual activity, including learning, decision-making, and natural language processing (Panesar, 2021). In particular, large language models (LLMs) – which can generate coherent texts and sustain dialogue – have been developing rapidly in recent years, leading to their widespread adoption (Ghassemi et al., 2023). This has become one of the reasons why large language models are increasingly being introduced into the humanities and educational sciences.

In psychology, AI is now being used with growing intensity. It should be noted that the orientation toward language models did not emerge suddenly. The precursors of "AI counseling" can be traced back to 1966, when Joseph Weizenbaum created the ELIZA program, which simulated a psychotherapeutic dialogue (Beg et al., 2024). Contemporary AI systems have significantly expanded these capabilities. They function as chatbots, virtual assistants, and specialized digital mental health platforms. Examples include Woebot and Tess, as well as Ukrainian platforms – Druh. Persha dopomoha, AI PSY HELP, and Faino – which are oriented toward psychological support and psychoeducation (Leonova, Leonov, 2025; AI PSY HELP; Faino). In Ukraine, these platforms have acquired particular relevance at the present time. The ongoing war has, first, heightened the population's need for psychological assistance and, second, has altered the familiar model of support. A psychologist is not always accessible (Dotsenko, 2025; Harmash, Pashchenko, 2024). For this reason, chatbots and specialized platforms often become the only channels through which rapid psychological help can be obtained (Leonova, Leonov, 2025).

A paradoxical situation is currently taking shape. On the one hand, we may speak of AI being used as a form of primary psychological assistance – that is, as a counselor. On the other hand, no consensus definition of "AI counseling" exists in the scientific literature. Therefore, in this paper we employ a working definition (which undoubtedly requires further operationalization and elaboration). AI counseling may be defined as a form of psychological assistance provided through AI systems. The substantive components of AI counseling include the analysis of a user's text messages and the provision of support based on the information obtained. Such support may take the form of delivering information the user requires, or of offering emotional support.

In addition, AI counseling may include a prospective component in the form of generating recommendations aimed at improving psychological well-being and self-regulation. It should be emphasized that research in this area confirms AI's capacity to conduct screening psychodiagnostics and, on that basis, to develop individualized psychological intervention plans (Ramirez, 2024). Furthermore, large language models are capable of analyzing the tone of a user's message and adapting their own messages to the user's linguistic behavior patterns (Zhao, 2023). Researchers emphasize that AI has high potential for expanding the accessibility of psychological assistance. This situation affects not only the expansion of opportunities for those seeking psychological help, but may also indirectly alter the employment structure of psychologists themselves. The use of AI counseling by psychologists allows them to focus on providing deeper forms of psychological intervention (Beg et al., 2024; Ramírez, 2024).

While acknowledging the growing significance of AI in psychological counseling, certain problems and challenges must also be noted. In particular, AI is capable of generating structured, stylistically polished texts that create an impression of depth and professionalism. Students may make use of such texts without attempting to engage with their actual context and content. The issue is that behind a seemingly substantive AI-generated text there may lie what could be called "contentless content". Recognizing this phenomenon requires a sufficient level of professional competence – one that students, unfortunately, do not yet possess to an adequate degree. As a result, they frequently perceive AI-generated texts as professional and expert-level, and incorporate them into their academic work. This situation creates yet another difficulty: students begin to uncritically accept AI-generated outputs as authoritative and expert.

Consequently, students may develop erroneous patterns of data interpretation and an excessive dependence on AI (Kim et al., 2025; Shen, Cui, 2024; Xia et al., 2025).

Thus, one of the key competencies in psychological training is the capacity for critical evaluation of AI-generated professional content. This involves several distinct operations: distinguishing description from interpretation, grounding conclusions in empirical data, and avoiding professionally unwarranted inferences. Before such competency can be developed, however, it is necessary to map the differences between students' interpretations of psychological material and those produced by AI.

This article therefore pursues a dual aim. The first is a theoretical analysis of AI's potential and limitations in psychological counseling. The second is an empirical one: we present the results of a pilot pedagogical experiment comparing psychology students' interview interpretations with those of AI, with the broader goal of identifying risks and benefits of integrating AI tools into psychological education.

2. Materials and methods

This study used a mixed-method design, including a theoretical literature review and an empirical pedagogical experiment. The theoretical component included a review of theoretical concepts and empirical research on the use of AI in psychological counseling. Specifically, interaction mechanisms were analyzed, including how AI simulates empathy, the cognitive-behavioral methods it uses, and how it structures dialogue. The limitations of AI were also explored, particularly in interactions with users experiencing complex emotional states or in crisis situations. The theoretical review yielded five key themes concerning the practical implications of AI integration in counseling: (1) adherence to psychotherapeutic methodology and protocols; (2) emotional support and creation of a psychologically safe environment; (3) formation of a therapeutic alliance; (4) ethical considerations and confidentiality; and (5) accessibility of psychological assistance using AI technologies. To empirically assess the differences between student and AI interpretations of interviews, a pilot pedagogical experiment was conducted. This experiment was integrated into the curricula of the courses "Experimental Psychology" (3rd-year, bachelor's degree), "Qualitative Research Methods" (4th-year, bachelor's degree), and "Organization and Methods of Scientific Research in Psychology" (1st-year, master's degree).

2.1. Participants and Procedure

This study employed a mixed-methods approach, combining a theoretical literature review with an empirical pedagogical experiment. The theoretical component involved examining conceptual frameworks and existing research on AI applications in psychological counseling. Specifically, we looked into interaction mechanisms, such as how AI models empathy, its use of cognitive-behavioral techniques, and the way it structures dialogue. Furthermore, the analysis touched upon the inherent limitations of AI, particularly when interacting with users experiencing complex emotional states or crisis situations.

Based on this review, we identified five key areas relevant to the practical implementation of AI in counseling: (1) adherence to psychotherapeutic methodology and protocols; (2) provision of emotional support and a psychologically safe environment; (3) the formation of a therapeutic alliance; (4) ethical considerations and confidentiality; and (5) the accessibility of mental health services through AI technologies.

To empirically evaluate the differences between interview interpretations made by students versus AI, we conducted a pilot pedagogical experiment. This experiment was integrated into the curricula of several courses: "Experimental Psychology" (3rd-year undergraduate), "Qualitative Research Methods" (4th-year undergraduate), and "Organization and Methods of Scientific Research in Psychology" (1st-year graduate).

The study was carried out in three stages, involving psychology students at different levels of their education. The data collection phase was handled by third-year students (N = 22; aged 19–21). For the subsequent analysis phase, we involved fourth-year undergraduate and first-year graduate students (N = 44; aged 21–23). All participants possessed foundational training in psychodiagnostics, qualitative research methods, and counseling principles.

Stage 1. Empirical Data Collection

At this stage, junior undergraduate students (N = 22) participated by developing semi-structured interview guides and interviewing respondents from their immediate social environment. The interview topics were everyday and non-clinical in nature (e.g., "Attitudes

toward pets", "Experience of making important decisions", "The role of friendship"). This focus was determined by the need to avoid ethical risks associated with working with clinical material. The interviews were subsequently transcribed and underwent preliminary technical text review. Each student produced a full verbatim transcript, resulting in a dataset of 22 complete transcripts used for subsequent comparative analysis.

Stage 2. Formulation of Analytical Tasks

Analytical questions requiring deeper interpretive engagement with the material were formulated for each transcript. Examples of such questions include: "What are the main motives underlying this student's academic activity?", "What influenced this respondent's choice of profession?", "What factors are primary in this respondent's decision-making in difficult situations?", and "How does the respondent assess their need for social support?" It should be noted that the questions were formulated on the basis of the specific content of each individual interview.

Stage 3. Parallel Analysis by Students and AI

A total of 22 text packages were prepared, each consisting of an interview transcript and the corresponding analytical questions. Each package was analyzed both by senior students and by an AI system.

(a) Forty-four senior students participated in the analysis of the text packages (4th-year undergraduates and 1st-year master's students). Each text package was independently analyzed by two students, yielding 44 analytical texts.

(b) The same text packages were analyzed by an AI system. Claude 3.5 Sonnet (Anthropic) was used for this purpose.

The resulting dataset comprised 66 interpretations in total (44 student-produced and 22 AI-generated). It should be emphasized that the dataset (44 student and 22 AI interpretations) was derived from the analysis of identical source material – a prerequisite for enabling meaningful comparison of interpretive strategies between humans and AI.

2.2. Text Corpora Characteristics

The student corpus totaled 21,520 words (M = 489.09 words per text), while the AI corpus comprised 15,922 words (M = 723.73 words per text). Lexical density (the ratio of unique words to total text volume) was comparable across both sets: 0.64 for student texts and 0.62 for AI texts. The readability index was 18.28 for the student corpus and 17.7 for the AI corpus. Notably, the average sentence length in the student texts was 16.18 words, whereas AI-generated sentences were substantially longer, averaging 27.98 words.

2.3. Data Analysis Procedures

The texts underwent two modes of analysis.

1. Marker-Based Discourse Analysis

To ensure maximum objectivity, we conducted a semi-automated discourse analysis of marker frequency. Prior to the analysis, the following markers were identified:

(a) Uncertainty markers (e.g., "probably", "one might assume", "it seems");

(b) Causal markers (e.g., "because/since", "therefore", "this indicates," "this suggests");

(c) Certainty markers (e.g., "obviously," "undoubtedly," "a key factor," "certainly," "it is clear that");

(d) Recommendation markers (e.g., "it is worth", "it is advisable", "it is recommended", "it is necessary").

The texts were reviewed manually in MS Word. Each text was searched for marker words and their occurrences were counted. Given that the resulting text corpora varied in length, the frequency of marker words and phrases was normalized per 1,000 words. All identified marker instances were additionally verified in context to exclude false positives (e.g., atypical usage).

We define over-interpretation as an analytical error in which a conclusion lacks sufficient empirical justification within the source material. We have identified three types of over-interpretation:

Type A. Conclusions drawn from an absence of information (e.g., concluding that a respondent prefers to spend time alone simply because the interview did not mention specific hobbies or social activities).

Type B. Quasi-diagnostic interpretations lacking sufficient empirical evidence or formal diagnostic validation (e.g., the use of terms such as "introvert", "high anxiety", or "suicidal ideation").

Type C. Overgeneralizations utilizing absolute terminology (e.g., "always," "never," "everyone", "certainly").

To eliminate the risk of subjectivity, the assessment of over-interpretation was conducted by three independent experts. These experts were psychologists experienced in the qualitative analysis of interviews. Each expert evaluated the presence or absence of over-interpretation within the texts. Identified cases were assessed based on pre-defined criteria and classified as Type A, B, or C.

The experts coded the texts independently. Overall, inter-rater reliability regarding the presence of over-interpretation was high (85–90 %). Discrepancies (10–15 %) most frequently occurred in borderline cases, were subtle, or reflected a combination of multiple types. These discrepancies were resolved through consensus-based discussion.

3. Discussion

A theoretical analysis of the literature on the capabilities and limitations of AI in psychological counselling allows us to draw several important observations. On one hand, AI technologies expand access to basic psychological support. At the same time, they are considerably limited when it comes to deep therapeutic contact or the need for nuanced empathic engagement with a client.

One can fully agree with researchers who emphasise the high effectiveness of AI when using algorithmised structured protocols – in particular within the cognitive-behavioural framework (Husnain et al., 2024; Jiang et al., 2024). Large language models can effectively identify cognitive distortions, are capable of sustaining logical dialogue, and can propose structured interventions. In other words, AI is quite effective when therapeutic influence is grounded in clear algorithms.

Psychology is currently developing a rather high regard for algorithmised intervention, which is viewed as a rapid and effective therapeutic approach. However, one can agree with Podolan and Gelo (2024), who argue that the outcome of psychotherapy depends not only on protocol adherence but, above all, on the competence of the practitioner and the effectiveness and quality of the therapeutic relationship. This view is reinforced by reports of AI's inflexibility.

Regarding emotional support provided by AI, a certain paradox is currently emerging. On one hand, users often perceive interaction with AI as emotionally safe due to its relative predictability and the absence of human evaluative judgements (Yahaya, Rukayya, 2024). On the other hand, AI frequently demonstrates "toxic positivity" and fails to grasp psychological ambivalence, pointing to the superficiality of algorithmised dialogue. At present, AI cannot provide deep empathic acceptance or process complex emotions, which renders interaction with it rather shallow (Peluso, Freund, 2018).

Thus, the formation of a genuine therapeutic alliance with AI currently appears to be fundamentally impossible. Despite the subjective feelings of support reported by some users, empirical data confirm that algorithm-based interaction with AI is qualitatively different from a human therapeutic alliance (Wang et al., 2025). AI does not possess the capacity for genuine mutuality, sincere self-disclosure, or reciprocal influence – the very foundations upon which real therapeutic relationships are built.

AI-assisted counselling also gives rise to ethical dilemmas that cannot currently be resolved within the framework of traditional clinical practice. The widespread adoption of AI technologies raises new questions concerning confidentiality, informed consent, and accountability. In interactions with AI, users may encounter the phenomenon of "diffused responsibility", whereby harm cannot be compensated or further assistance obtained (Jiang et al., 2024; Ramírez, 2024). It should also be noted that behind AI's apparent democratisation lies digital inequality: access to AI systems may be restricted for certain social groups due to a lack of devices, internet connectivity, or electricity.

One argument in favour of AI use may be comprehensive technological support from governmental or non-governmental structures, particularly in force majeure or extreme circumstances. Location-independent support genuinely expands access to basic psychiatric care, especially in crisis situations such as the ongoing war in Ukraine (Leonova, Leonov, 2025).

Within our pilot pedagogical project, we sought to examine how certain characteristics of AI manifest within an educational paradigm. A comparison of interview interpretations produced by students and by artificial intelligence revealed consistent divergences in analytical style and distinct risks of interpretive error. For instance, student texts more frequently contained markers of caution, reflecting a tendency toward careful consideration of assumptions and acknowledgement of the limits of one's own experience. Texts generated by AI more frequently employed markers of causal reasoning and recommendation. This linguistic profile superficially

resembled an authoritative "expert" style, yet was accompanied by over-interpretation, making the AI's conclusions excessively categorical.

Notably, categorical conclusions were systematically drawn on the basis of absent information. For example, AI interpreted the absence of any mention of a particular life domain in an interview as conclusive evidence of its insignificance in the respondent's life. This empirical finding underscores the need to reconceptualise the role of AI in psychological education. AI output should not be treated as a standard of professional psychological interpretation; rather, AI may be used as a supplementary educational tool through which students can learn to recognise hidden analytical errors. This pedagogical strategy should, on one hand, employ AI as an assistive instrument for working with texts, and on the other, train students to critically evaluate both texts and their overall engagement with AI.

4. Results

4.1. Systematic Review Results: Key Dimensions of AI Counseling

A theoretical analysis of the literature on the assessment of AI's capabilities in counselling identified a number of key aspects. AI systems are primarily able to follow protocols and maintain the structure of psychotherapeutic interaction. While this may appear to be a condition for therapeutic effectiveness, strict protocol adherence in psychotherapy does not always guarantee a therapeutic outcome. The study by Podolan and Gelo revealed a weak correlation between protocol adherence and clinical outcomes, leading the authors to conclude that adherence to protocol is an important but not absolute condition for effective therapeutic intervention. In other words, AI counselling algorithms perform well within structured methods but are less predictable in situations of high uncertainty.

Contemporary applications such as Woebot and Tess have well-refined algorithms that enable them to effectively simulate session sequences, sustain engagement, and monitor users' emotional states. The most compelling results in this context are produced by structured approaches – most notably cognitive behavioural therapy (CBT). Jiang et al. demonstrated that modern chatbots employ natural language processing (NLP) and large language models (LLMs) to identify cognitive distortions, reproduce the stages of cognitive restructuring, generate response scenarios, and track respondents' progress. On the other hand, in less formalised therapeutic approaches, AI demonstrates a number of limitations. Wang et al. showed that, despite a natural conversational interface, chatbots lack the flexibility required for interpersonal interaction. Thus, AI may be quite effective in interventions grounded in algorithms, but has limited effectiveness in non-formalised interaction.

4.1.2. Emotional Support and the Creation of a Safe Environment

The primary indicators of effective AI-assisted counseling, according to users, are emotional support and the subjective perception of psychological safety. At the same time, the concept of safety is understood differently in the context of interaction with a human counselor versus an AI counselor.

Safety within traditional psychotherapy relies on a secure environment for self-exploration and processing of complex emotions (Podolan, Gelo, 2024), whereas AI counseling reconceptualizes this dynamic. Safety in the context of AI interaction is grounded in predictability, the absence of human evaluative judgments, and a degree of anonymity. The study by Yahaya and Ruqayyah confirmed that users frequently perceive AI counselors as safe, given that interactions with them involve no negative assessments and, as a result, users do not experience stigmatization. An additional safety-related factor is the stability of algorithmic responses. Wang et al. (2025) emphasized that LLM-based chatbots demonstrate response consistency that is free from human emotional fluctuations, thereby enhancing the user's sense of control and predictability.

A limitation of AI algorithms is "toxic positivity," whereby the algorithm produces optimistic responses and comments regardless of context, which may invalidate the user's experience. Furthermore, the inability of AI to process nonverbal cues results in a simplified and superficial interpretation of the client's emotional state (Wang et al., 2025).

4.1.3. Formation of the Therapeutic Alliance

The criteria for a therapeutic alliance include emotional connection, goal alignment, and shared task performance (Podolan, Gelo, 2024; Prusiński, 2022; Tschuschke et al., 2020). While these predictors are reliable indicators of effectiveness in human-human interactions, many questions remain in human-AI interactions. A fundamental question is whether AI can form a

high-quality therapeutic alliance that goes beyond simply providing structured support. A review of the literature revealed two main trends.

On the one hand, users often perceive chatbots as supportive partners due to their constant availability, non-judgmental stance, and emotional predictability. This communication style reduces user anxiety and may facilitate self-disclosure. Such AI-based tools are particularly valuable for individuals who avoid interacting with human professionals due to shame or fear of stigma (Beg et al., 2024; Seow et al., 2021; Yahaya, Rukayya, 2024). On the other hand, despite using empathically coded language, AI cannot create the depth of relationship characteristic of human therapeutic contact; that is, interactions with AI are more functional than a mutually evolving alliance (Wang et al., 2025).

The inability to achieve genuine empathic reciprocity is a critical limitation of AI. Algorithms can mimic the linguistic markers of empathy, although they actually lack lived experience. The lack of real human experience and emotional depth limits interactions with users and undermines the mechanisms associated with authenticity, rupture-repair cycles, and relationship deepening (Wang et al., 2025). Although AI can create the illusion of support, which is necessary for both initial contact and consolidating gains between sessions (Beg et al., 2024; Yahaya, Rukayya, 2024), its ability to build a comprehensive therapeutic alliance remains limited (Podolan, Gelo, 2024; Wang et al., 2025). It is therefore crucial that students understand these features of AI algorithms. Psychological education should clearly define this boundary: AI functions as a tool, but cannot replace human interaction as the foundation of the therapeutic alliance.

4.1.4. Ethical Risks and Confidentiality

The integration of AI-assisted counseling into practice gives rise to a number of ethical dilemmas that extend beyond traditional therapeutic boundaries. While practicing professionals operate within strict professional guidelines and codes of ethics, AI systems currently function within ambiguous regulatory frameworks. This creates risks related to confidentiality, algorithmic transparency, accountability, and the scalability of potential harm (Beg et al., 2024). Privacy concerns are of particular importance, especially with regard to minors, who frequently turn to AI counselors with their questions – a context in which there is no means of monitoring the level of privacy protection or ensuring informed consent (Ardity, Thompson, 2023; Kafka et al., 2024).

Data security is of critical importance. AI platforms collect vast amounts of sensitive information, including dialogue content and metadata – such as frequency of use and emotional trajectories – stored within digital ecosystems (Ramírez, 2024). This situation is considerably vulnerable, further compounded by opaque privacy policies that remain largely incomprehensible to users (Alfano et al., 2023; Beg et al., 2024; Fiske et al., 2019). Informed consent may be further complicated by the anthropomorphization of AI. Users frequently overestimate the capabilities of chatbots, projecting human-like understanding onto algorithmic responses (Eberle et al., 2021; Wang et al., 2025). This leads to excessive trust in AI-generated information, potentially guiding users toward erroneous decisions (Gipps, 2023; Wang et al., 2025).

The distribution of responsibility also presents a serious challenge. In AI-assisted counseling, accountability is dispersed among developers, platform providers, and implementing organizations. This creates a regulatory vacuum regarding liability for algorithm-induced harm – a risk that is further amplified by the scalability of the technology (Jiang et al., 2024). Additionally, algorithmic bias and cultural adequacy require careful scrutiny. Models trained predominantly on WEIRD datasets (Western, Educated, Industrialized, Rich, and Democratic) may generate recommendations that are ill-suited to the cultural contexts of other social groups (Ramírez, 2024). Furthermore, the digital divide limits genuine accessibility for vulnerable populations, contradicting the technology's claims of democratization. Consequently, AI-related ethical competence – encompassing critical evaluation of risks concerning privacy, bias, and accountability – must become a mandatory component of professional training for psychologists (Beg et al., 2024; Hunt, Blease, 2024; Jiang et al., 2024; Ramírez, 2024; Wang et al., 2025).

4.1.5. Accessibility of Psychological Assistance

One of the key advantages of AI counseling is the expansion of accessibility to psychological support. Artificial intelligence technologies are capable of partially removing the geographical, temporal, economic, and social barriers inherent in traditional therapy (Aljunaidel et al., 2024). However, this expansion of accessibility is uneven and is accompanied by the emergence of new risks.

AI is able to overcome geographical barriers, which is particularly significant in crisis contexts. Leonova and Leonov highlighted the importance of AI-based mental health tools in the

context of the war in Ukraine, where traditional infrastructure has been disrupted and populations in conflict zones lack access to in-person assistance. Local digital initiatives – such as Druh. Persha dopomoha, AI PSY HELP, and Faino – have already demonstrated the applied potential of such solutions in situations where conventional psychological support is unavailable. Temporal accessibility is also substantially enhanced: AI-based services provide round-the-clock support (24/7) and enable immediate intervention during acute distress (Husnain et al., 2024). These features reduce clients' dependence on specialists' schedules and make interaction with an AI counselor more flexible. In addition, AI improves the economic accessibility of psychological support through free or subscription-based models (Ramírez, 2024). However, the monetization of accessibility is frequently accompanied by the collection of user data, resulting in a violation of the confidentiality principle. In effect, users pay for platform access at the cost of their own anonymity (Olawade et al., 2024; Ramírez, 2024).

Importantly, AI also lowers the psychological barriers to help-seeking. Perceived anonymity and the absence of judgment make AI an attractive point of first contact for users who avoid traditional therapy due to stigma (Yahaya, Rukayya, 2024). Nevertheless, technological anonymity does not guarantee genuine confidentiality, and using AI to circumvent stigma does not address the social roots of its existence (Beg et al., 2024; Yahaya, Rukayya, 2024).

While multilingual AI models broaden access for linguistic minorities, Jiang et al. note that linguistic accessibility is not equivalent to cultural appropriateness. AI counseling models require adaptation to the cultural characteristics of their users – in particular, to their perceptions of mental health and their ways of expressing distress. On one hand, AI may be especially valuable for certain social groups, such as individuals with social anxiety or limited mobility; on the other hand, it may be inaccessible or potentially unsafe for groups with different characteristics – for example, older adults or individuals with cognitive impairments, who may be classified as having a high degree of digital vulnerability.

Based on the foregoing, a hybrid interaction model can be proposed that combines elements of AI counseling with traditional psychological support. In such models, AI provides initial support, psychoeducation, and symptom monitoring, while human professionals assume responsibility for crisis intervention, complex casework, and the development of therapeutic relationships (Beg et al., 2024). Implementing this approach requires the development of clear protocols to delineate the respective responsibilities of humans and AI, as well as to identify situations in which human intervention is necessary (Beg et al., 2024; Jiang et al., 2024).

Despite the genuine expansion of accessibility that AI offers, there is a risk of fostering a two-tier system in which AI replaces human therapy for socially vulnerable populations (Beg et al., 2024; Ramírez, 2024). Professional training should therefore emphasize that technological accessibility is not equivalent to therapeutic effectiveness. Future psychologists must be prepared for AI to serve as the first point of contact in the help-seeking process – particularly in crisis or resource-limited situations. This calls for the development of competencies in evaluating AI-generated content, managing processes that involve AI, and effectively routing clients between algorithmic and human resources.

4.2. Empirical Results: Pilot Pedagogical Experiment

4.2.1. Marker Analysis and Overinterpretation Indicators

An analysis of markers and independent assessments of overinterpretation revealed systematic stylistic discrepancies between the student and AI interpretations. Student responses and AI texts represented two qualitatively different analytical profiles. Student texts contained significantly higher frequencies of markers of reflexivity and epistemic caution (e.g., "in my opinion," "one can assume," "probably"). This reflected a tendency to carefully formulate assumptions about the causes of respondents' behavior and avoid premature generalizations. From a pedagogical perspective, such an analysis indicates professional rigor and well-developed critical thinking, as evidenced by students' ability to distinguish facts from interpretations and recognize the limitations of the data.

On the other hand, AI responses more frequently contained markers of causality and structural argumentation (e.g., "since," "therefore," "this indicates," "leads to"). Such constructions created the impression of a logically coherent and systematic analysis. Furthermore, AI-generated texts contained significantly more frequent markers of recommendations and interventions (e.g., "worth," "recommended"), along with lexemes related to "strategies" and "methods," emphasizing a pronounced

focus on practical and actionable results. The AI-generated corpus also more frequently contained markers of confidence (e.g., "obviously," "undoubtedly," "clearly indicates").

Overall, the AI style conveyed an impression of expertise and competence. However, upon closer analysis, AI texts appeared superficial, inattentive to contextual nuances and factual detail, and reflected an inability to work with incomplete data. Quantitative analysis revealed that student texts contained more than three times as many epistemic caution markers as AI texts (3.84 vs. 1.20 per 1,000 words), while AI-generated texts contained more than twice as many causal inference markers (4.58 vs. 1.92) and recommendation markers (3.61 vs. 1.60). Table 1 presents summary marker frequencies. This linguistic methodology is consistent with research demonstrating that AI-generated texts differ from human writing in their marker use and syntactic patterns (Botes et al., 2025). This study adapted this approach to evaluate interpretive strategies in psychology education.

Table 1. Frequency of Discourse Markers in Student and AI Texts (per 1,000 words)

Category	Students (Count)	Students (per 1,000 words)	AI (Count)	AI (per 1,000 words)
Hedging markers (epistemic uncertainty)	83	3.86	19	1.19
Evidentiary markers (data grounding)	58	2.70	38	2.39
Assertiveness markers	10	0.46	11	0.69
Causality/reasoning markers	41	1.90	73	4.59
Ambivalence/contrast markers	107	4.97	77	4.84
Recommendation/intervention markers	34	1.58	57	3.58

Notes: Frequencies are normalized per 1,000 words to account for corpus volume disparities (Student corpus = 21,520 words; AI corpus = 15,922 words).

4.2.2. Overinterpretation as a Key Didactic Risk

Identifying overinterpretation – conclusions lacking sufficient empirical grounding – constituted a critical analytical vector. This error type poses a severe pedagogical risk because it is frequently masked by academic formatting, logical structure, and authoritative phrasing, generating an illusion of evidence-based professionalism. The analysis indicated that AI texts most frequently manifested Type A overinterpretation (conclusions derived from information absence). The AI systematically interpreted the absence of a direct mention in the interview as sufficient grounds for categorical conclusions regarding a trait's presence or absence. For instance, if family support was unmentioned, the AI concluded the respondent was socially isolated.

This interpretational logic is methodologically invalid: an omission may reflect the conversation's specific focus, protocol limitations, or respondent reticence, and cannot automatically denote the phenomenon's absence. This is pedagogically hazardous as it establishes an erroneous standard for data interpretation. Observing AI routinely draw confident conclusions from absent data risks students internalizing this flawed logic as normative. Professional psychological analysis requires strictly differentiating established facts, justified interpretations, and speculative assumptions requiring further data.

Overinterpretation occurred less frequently in student texts and was generally mitigated by modal markers of supposition (e.g., "probably," "one might assume"). However, students also exhibited instances of unjustified categorization (using "obviously") and deployed diagnostic constructs requiring specialized assessment. Overall, AI texts demonstrated a stronger propensity for categorical generalizations based on limited or ambiguous data. Specifically, the AI routinely interpreted omissions as indicators of insignificance – a profound methodological error. Conversely, students navigating incomplete information maintained a more rigorous professional stance by deploying supposition markers.

Consequently, AI should not serve as a standard for correct interpretation. Instead, it operates as educational material enabling students to develop critical analysis skills: identifying hidden interpretational errors, deconstructing rhetorical strategies that manufacture artificial

evidence, and delineating the boundaries of permissible clinical conclusions. Tables 2 and 3 present the quantitative frequencies and illustrative examples of overinterpretation types.

Table 2. Frequency of Overinterpretation Types in Student and AI Texts (per 1,000 words)

Type of Overinterpretation	Students (Count)	Students (per 1,000 words)	AI (Count)	AI (per 1,000 words)
A: Conclusions from absence of information	3	0.14	4	0.25
B: Diagnostic/pathologizing language	21	0.98	23	1.44
C: Excessive generalizations	14	0.65	27	1.70

Table 3. Illustrative Examples of Overinterpretation in Student and AI Texts

Type of Overinterpretation	Student Example	AI Example
A: Conclusions from absence of information	"Close ones and the social environment played an important role, but in the respondent's answers... friends and support are not mentioned, which may indicate her less active participation."	"Family is not directly mentioned – probably, it is secondary in this situation."
B: Diagnostic/pathologizing language	"The respondent is exhausted: experienced stress [related to fatigue?], and this led to procrastination."	"During the most difficult period (second session in 1st year) she experienced apathy and devastation."
C: Excessive generalizations	"After encountering difficulties during university studies, the girl finds support in her own experience... sometimes late procrastination can be a way for her to confirm her capability."	"Communication with classmates almost does not occur; as a result, some of them stopped studying or were mobilized."

4.2.3. Summary of Results and Pedagogical Conclusions

Overall, AI demonstrates strong performance in analyzing text structure and drawing logically grounded conclusions. AI algorithms effectively systematize material, identify thematic blocks, and construct cause-and-effect relationships. Student texts, by contrast, are more reflexive, epistemically cautious, and reflect an awareness of data limitations. However, student texts frequently lack the compositional coherence characteristic of AI-generated output – a situation that most likely reflects the process of developing professional identity, as students learn to balance interpretive caution with analytical confidence.

The "incompleteness" of student analyses creates a pedagogical vulnerability: struggling to produce compositionally coherent texts, students may come to regard AI-generated texts as a professional standard. The idealization of AI output and its uncritical adoption exacerbates categorical and overreaching interpretation among students. AI systems effectively create an illusion of analytical completeness by generating pseudo-scientific terminology, while in reality their texts are often fragmentary or ambiguous. This apparent "pseudo-expertise" of AI promotes uncritical acceptance of generated content and may instill flawed interpretive paradigms in students.

The experimental findings support several practical principles for integrating AI tools into psychology education. Central to this integration is a reconceptualization of AI's role within the curriculum. One of the key professional competencies of a psychologist is the ability to distinguish

between levels of evidential support for clinical claims. This means that students must learn to rigorously differentiate between established facts ("the respondent reported..."), justified interpretations ("this may indicate..."), and hypothetical assumptions ("probably," "one might suggest"). Instructors should emphasize that the use of epistemic caution markers reflects professional maturity and epistemological rigor – not analytical weakness.

This reflects the principle of interpretive evidence: every substantive claim must be grounded in specific empirical data, such as direct quotations, observable behavior, or psychodiagnostic findings. Accordingly, learning assignments should be designed to develop students' ability to justify their key professional conclusions, as this constitutes an essential condition for the development of psychological thinking.

An important task for instructors is the systematic identification and critical analysis of overgeneralizations, inferences drawn from absent information, and unsubstantiated quasi-diagnostic formulations in student texts. Concrete examples of interpretations can also be used as didactic material in group discussions. Comparative analysis of student texts and AI-generated texts may serve as an effective mechanism for developing professional reflection. Assignments requiring students to compare their own texts with AI output, identify typical algorithmic errors, and deconstruct concealed rhetorical strategies will effectively foster critical evaluation skills, preparing future psychologists for the responsible clinical use of AI. The integration of AI into psychology education thus implies the formation of a professional stance that consciously defines both the capabilities and the limitations of AI algorithms. It is therefore essential that, in the course of their training, students learn to critically evaluate AI-generated content and understand which clinical tasks can be safely delegated to an algorithm.

Limitations

This pilot study acknowledges several limitations. First, the restricted sample size (N = 22 interviews; 44 students and 22 AI interpretations) constrains statistical power and the generalizability of the findings. Second, the interviews focused on everyday, non-clinical topics; analyzing clinical material may yield different interpretational dynamics. Third, the students were aware of their participation in the study, potentially inducing a Hawthorne effect that increased their linguistic caution. However, the identified discrepancies remain systematic and quantitatively robust, exhibiting two- and threefold differences in key discourse markers. The students' threefold predominance in epistemic caution (3.84 vs. 1.20 per 1,000 words) and the AI's twofold predominance in causality markers (4.58 vs. 1.92) transcend random variance, reflecting fundamentally divergent analytical strategies.

Prospects for Further Research

These preliminary findings establish several trajectories for future research. First, longitudinal studies are necessary to determine if systematic exercises in critically analyzing AI content cultivate enduring skills in recognizing overinterpretation and ultimately enhance graduates' clinical proficiency. Second, future research should incorporate clinical datasets to evaluate whether the observed human-AI discrepancies persist in complex diagnostic scenarios requiring differential diagnosis and multilevel interpretation. Third, dedicated empirical evaluation of Ukrainian AI psychological support platforms (e.g., AI PSY HELP, Faino, Druh) is critical. Assessing their clinical efficacy, cultural adequacy, and impact on help-seeking behaviors is essential for establishing an evidence base for their integration into national mental health infrastructure. Finally, a comparative analysis of various AI models (e.g., Claude, ChatGPT, specialized clinical chatbots) regarding the frequency and typology of interpretational errors is required to identify the safest systems for educational deployment.

5. Conclusion

This study demonstrates qualitative discrepancies in the interpretive strategies employed by psychology students and AI systems when analyzing qualitative interviews. Student texts are characterized by pronounced epistemological caution and reflexivity. AI-generated texts, by contrast, exhibit formal structural coherence and linguistic markers of confidence. At the same time, AI texts are oriented toward overinterpretation, drawing categorical conclusions without sufficient empirical grounding in the source material.

The literature review confirms the considerable potential of AI-assisted counseling – particularly in democratizing access to basic psychological support, especially in crisis situations. However, the clinical effectiveness of AI remains structurally limited, given its weak capacity for

processing deep emotional experiences, forming a therapeutic alliance, and managing the nuanced interpersonal dimensions of clinical contact. The most viable operational model may be the integration of AI as a supplement to traditional psychological support, rather than its deployment as an autonomous clinical unit.

In professional psychological education, the primary goal should shift from simply training students to use AI tools toward developing the ability to critically evaluate AI-generated content. The constituent elements of such training may include learning to identify concealed interpretive errors, rigorously upholding evidential standards in the formation of clinical conclusions, and maintaining the epistemological boundaries of psychological interpretation. With targeted pedagogical support, AI can serve as a highly effective didactic resource for developing students' critical thinking – provided it is not treated as a standard of analytical interpretation, but rather as a dynamic resource for identifying and mitigating typical diagnostic risks. Ultimately, the integration of AI into clinical training requires the formation of a new professional competency: the capacity to accurately assess the limitations of algorithms, establish parameters for their clinical application, and actively preserve the therapeutic presence of the human practitioner as the irreplaceable foundation of psychological support.

References

Alfano et al., 2023 – *Alfano, L., Malcotti, I., Ciliberti, R.* (2023). Psychotherapy, artificial intelligence and adolescents: Ethical aspects. *Journal of Preventive Medicine and Hygiene*. 64: E438-E442.

Aljunaidel et al., 2024 – *Aljunaidel, N., Albattal, S., Kofi, M.* (2024). Measuring the impact of and challenges in using psychotherapy sessions in primary health care, Riyadh, Saudi Arabia. *European Journal of Medical and Health Research*. 2: 98-103.

Ardity, Thompson, 2023 – *Ardity, R., Thompson, P.* (2023). Effects of client confidentiality on adolescents' willingness to attend therapy. *Journal of Student Research*. 12.

Beg et al., 2024 – *Beg, M.J., Verma, M.* (2024). Artificial intelligence for psychotherapy: A review of the current state and future directions. *Indian Journal of Psychological Medicine*. DOI: 10.1177/02537176241260819

Botes et al., 2025 – *Botes, E., Dewaele, J.-M., Colling, J., Teuber, Z.* (2025). Initial indications of generative AI writing in linguistics research publications [Preprint]. *OSF*. DOI: 10.31234/osf.io/4yvbp_v1

Dotsenko, 2025 – *Dotsenko, V.* (2025). Psykholohichna dopomoha v umovakh viiny: Vyklyky ta shliakhy podolannia [Psychological assistance in wartime: Challenges and ways to overcome them]. *Problemy psykholohii diialnosti v osoblyvykh umovakh: Materialy III Vseukr. nauk.-prakt. konf.* Pp. 223-224. [in Ukrainian]

Eberle et al., 2021 – *Eberle, K., Grosse Holtforth, M., Inderbinen, M., Gaab, J., Nestoriuc, Y., Trachsel, M.* (2021). Informed consent in psychotherapy: A survey on attitudes among psychotherapists in Switzerland. *BMC Medical Ethics*. 22: 150.

Fiske et al., 2019 – *Fiske, A., Henningsen, P., Buyx, A.* (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*. 21.

Ghassemi et al., 2023 – *Ghassemi, M., Birhane, A., Bilal, M., Kankaria, S., Malone, C., Mollick, E., Tustumi, F.* (2023). ChatGPT one year on: Who is using it, how and why? *Nature*. 624: 39-41.

Gipps, 2023 – *Gipps, R. G. T.* (2023). Psychotherapy as ethics. *Philosophies*. 8: 42.

Harmash, Pashchenko, 2024 – *Harmash, A.O., Pashchenko, Ye.A.* (2024). Psykholohichnyi vplyv viiny na osobystist, pidkhody ta stratehii dopomohy [Psychological impact of war on personality, approaches and strategies of assistance]. *Tsilisnyi pidkhid u psykholohii ta sotsialnii roboti: Teoriia ta praktyka*. Pp. 231-233. [in Ukrainian]

Hunt, Blease, 2024 – *Hunt, J., Blease, C.* (2024). Re-visiting professional ethics in psychotherapy: Reflections on the use of talking therapies as a supportive adjunct for myalgia encephalomyelitis/chronic fatigue syndrome and 'medically unexplained symptoms. *Journal of Medical Ethics*. DOI: 10.1136/jme-2023-109627

Husnain et al., 2024 – *Husnain, A., Ahmad, A., Saeed, A., Din, S.M.U.* (2024). Harnessing AI in depression therapy: Integrating technology with traditional approaches. *International Journal of Science and Research Archive*. 12: 2585-2590.

Jiang et al., 2024 – *Jiang, M., Zhao, Q., Li, J., Wang, S., He, T., Cheng, X., Yang, B.X., Ho, G.W.K., Fu, G.* (2024). A generic review of integrating artificial intelligence in cognitive behavioral therapy. *arXiv*. DOI: 10.48550/arxiv.2407.19422

Kafka et al., 2024 – *Kafka, J., Kothgassner, O.D., Felnhofer, A.* (2024). A matter of trust: Confidentiality in therapeutic relationships during psychological and medical treatment in children and adolescents with mental disorders. *Journal of Clinical Medicine*. 13: 1752.

Karimzadeh, Saeedi, 2025 – *Karimzadeh, D., Saeedi, A.* (2025). AI for mental health assessment and intervention: A systematic review. *International Journal of Modern Achievement in Science Engineering and Technology*. 5: 96-104.

Kim et al., 2025 – *Kim, Y., Kim, J.H., Zhang, S.* (2025). Trusting AI too much? Psychological predictors of overtrust and the mitigating role of AI literacy. *Research Square* (Preprint). DOI: 10.21203/rs.3.rs-7315296/v1

Lau et al., 2025 – *Lau, Y., Ang, W.H.D., Ang, W.W., Pang, P.C., Wong, S.H., Chan, K.S.* (2025). Artificial intelligence–based psychotherapeutic intervention on psychological outcomes: A meta-analysis and meta-regression. *Depression and Anxiety*. 2025: 8930012. DOI: 10.1155/da/8930012

Leonova, Leonov, 2025 – *Leonova, A.O., Leonov, M.A.* (2025). Tsyfrova psykholohichna pidtrymka uchasnykiv osvitnoho protsesu u voiennyi period [Digital psychological support for participants in the educational process during wartime]. *Materialy rehionalnoho naukovo-praktychnoho seminaru «Aktualni vektory rozvytku osvitnoi haluzi Ukrainy u voiennyi i pisliavoiennyi periody (do 95-richchia zasnuvannia Kryvorizkoho pedahohichnoho)»*. Pp. 103-106. [in Ukrainian]

Manole et al., 2024 – *Manole, A., Cârciumaru, R., Brînzaş, R., Manole, F.* (2024). An exploratory investigation of chatbot applications in anxiety management: A focus on personalized interventions. *Information*. 16: 11.

Olawade et al., 2024 – *Olawade, D., Wada, O.Z., Odetayo, A., David-Olawade, A.C., Asaolu, F.T., Eberhardt, J.* (2024). Enhancing mental health with artificial intelligence: Current trends and future prospects. *Global Medicine and Health*. 1: 100099. DOI: 10.1016/j.glmedi.2024.100099

Panesar, 2021 – *Panesar, A.* (2021). What is artificial intelligence? *Machine Learning and AI for Healthcare*. Berkeley: Apress. DOI: 10.1007/978-1-4842-6537-6_1

Peluso, Freund, 2018 – *Peluso, P.R., Freund, R.R.* (2018). Therapist and client emotional expression and psychotherapy outcomes: A meta-analysis. *Psychotherapy*. 55: 461-472.

Podolan, Gelo, 2024 – *Podolan, M., Gelo, O.C.G.* (2024). The role of safety in change-promoting therapeutic relationships: An integrative relational approach. *Clinical Neuropsychiatry*. 21: 403-417.

Prusiński, 2022 – *Prusiński, T.* (2022). The working alliance and the short-term and long-term effects of therapy: Identification and analysis of the effect of the therapeutic relationship on patients' quality of life. *Psychiatria Polska*. 56: 571-590.

Ramírez, 2024 – *Ramírez, L.* (2024). Artificial intelligence in psychological diagnosis and intervention. *LatIA Journal*. 1: 26.

Seow et al., 2021 – *Seow, L.S.E., Sambasivam, R., Chang, S., Subramaniam, M., Lu, H.S., Assudani, H.A., Tan, C.-Y.G., Vaingankar, J.A.* (2021). A qualitative approach to understanding the holistic experience of psychotherapy among clients. *Frontiers in Psychology*. 12: 667303. DOI: 10.3389/FPSYG.2021.667303

Shen, Cui, 2024 – *Shen, Y., Cui, W.* (2024). Perceived support and AI literacy: The mediating role of psychological needs satisfaction. *Frontiers in Psychology*. 15: 1415248. DOI: 10.3389/fpsyg.2024.1415248

Teixeira da Silva, Yamada, 2024 – *Teixeira da Silva, J.A., Yamada, Y.* (2024). Could generative artificial intelligence serve as a psychological counselor? Prospects and limitations. *Central Asian Journal of Medical Hypotheses and Ethics*. 5: 297-303.

Tschuschke et al., 2020 – *Tschuschke, V., Koemeda-Lutz, M., von Wyl, A., Crameri, A., Schulthess, P.* (2020). The impact of patients' and therapists' views of the therapeutic alliance on treatment outcome in psychotherapy. *Journal of Nervous and Mental Disease*. 208: 56-64.

Wang et al., 2025 – *Wang, Y., Wang, Y., Ye, X., Escamilla, L., Augustine, B., Crace, K., Zhou, G., Zhang, Y.* (2025). Evaluating an LLM-powered chatbot for cognitive restructuring: Insights from mental health professionals. *arXiv*. DOI: 10.48550/arxiv.2501.15599

Xia et al., 2025 – *Xia, Q., Zhang, P., Huang, W., Chiu, T.K.F.* (2025). The impact of generative AI on university students' learning outcomes via Bloom's taxonomy: A meta-analysis and pattern mining approach. *Asia Pacific Journal of Education*. 45: 1-31.

Yahaya, Rukayya, 2024 – *Yahaya, T., Rukayya, A.* (2024). Artificial intelligence in mental health counseling: History, innovations, ethics, and future prospects. *Artificial Intelligence (AI) for Sustainable Futures in Education and Research*.

Zhao, 2023 – *Zhao, S.* (2023). Psychological healing in the digital age: A study of personalized collaborative models empowered by GAI. *ACM International Conference Proceeding Series*. DOI: 10.1145/3644116.3644221

Zhou et al., 2022 – *Zhou, S., Zhao, J., Zhang, L.* (2022). Application of artificial intelligence on psychological interventions and diagnosis: An overview. *Frontiers in Psychiatry*. 13: 811665. DOI: 10.3389/fpsyt.2022.811665