Dahò and Caci *BMC Psychology* (2026) 14:25
https://doi.org/10.1186/s40359-025-03729-2

BMC Psychology

RESEARCH

Open Access

Exploring AI-assisted design of executive function rehabilitation programs for individuals with ADHD: a mixed-methods evaluation of prompts and chatgpt outputs

Margherita Dahò[1,2*] and Barbara Caci[1,2]

Abstract

Background As Artificial Intelligence (AI) tools like ChatGPT gain traction in clinical contexts, their role in neurorehabilitation, particularly in addressing executive function impairments associated with ADHD, remains underexplored. This study examines whether generative AI can meaningfully support clinicians in designing individualized cognitive rehabilitation plans, not as a replacement but as a complementary aid.

Methods The research consisted of three separate studies, each addressing distinct stages of the investigation. First, expert-driven prompts were developed based on literature and clinical insights to guide ChatGPT in generating rehabilitation plans for three hypothetical profiles of individuals with ADHD (adolescents, adults, and older adults). In Study 2, the outputs were analyzed using a semi-systematic qualitative framework (ISAAC), assessing structure, coherence, and adaptability across developmental stages. Study 3 involved an external panel of 27 neuropsychologists and cognitive rehabilitation specialists (M = 6; F = 21; mean age = 46.5, SD = 15) who rated each plan's theoretical validity, clinical relevance, and feasibility.

Results Experts in Study 3 generally responded positively to the theoretical consistency of the plans, especially those for adolescents and adults, recognizing alignment with established models of executive function rehabilitation. Many professionals expressed openness to using AI as a support tool in practice. However, feasibility emerged as a key limitation, with concerns over a lack of personalization, unrealistic resource assumptions, and unvalidated techniques, particularly in adult and older adult profiles. These findings align with earlier studies in occupational therapy and clinical decision-making, which also identified challenges in real-world applicability.

Conclusion While clinical experts express cautious optimism about AI-assisted rehabilitation planning, further development is necessary to enhance accuracy, personalization, and feasibility for the safe integration of AI into clinical practice.

Keywords ADHD, Artificial intelligence, ChatGPT, Cognition, Executive functions, Human-Computer interaction, Neuropsychological rehabilitation, Neuropsychology

*Correspondence:
Margherita Dahò
Margherita.daho@unipa.it

[1]Department of Psychology, Educational Science and Human Movement, University of Palermo, Viale delle Scienze, Ed.15, Palermo 90146, Italy
[2]WeSearch Lab - Laboratory of Behavioral Observation and Research on Human Development, University of Palermo, Palermo, Italy

Introduction

ADHD and Executive Functions

Attention-Deficit/Hyperactivity Disorder (ADHD) is a complex neurodevelopmental condition characterized by symptoms of inattention, hyperactivity, and impulsivity that manifest early in development and interfere with functioning across multiple domains [1]. While traditionally associated with childhood, ADHD is now recognized as a lifelong disorder, affecting approximately 7% of children, 5.6% of teenagers, and 2.5% –6% of adults globally [2–4]. A central but often underemphasized feature of ADHD is executive dysfunction. Individuals with ADHD frequently show deficits in core executive functions, including inhibitory control, working memory, cognitive flexibility, planning, and sustained attention [5, 6]. These functions are essential for regulating behavior, maintaining focus, and adapting to new or complex situations. Impairments in these areas can thus affect academic achievement, occupational performance, social relationships, daily life functioning, and general quality of life [7–12]. Quintero-Lopez et al., (2024). Such executive difficulties are not uniform and manifest differently across developmental stages [7, 13]. For instance, adolescents with ADHD may struggle with organization and impulse control, while adults may encounter challenges in managing responsibilities and long-term planning [9, 14, 15]. Older adults with ADHD, a population historically under-recognized, may experience accelerated cognitive decline or social isolation due to persistent executive challenges [10, 16]. Addressing executive function impairments is thus a crucial target for intervention across the lifespan.

Challenges in intervention planning

Given the diversity of ADHD presentations and the way symptoms evolve across the lifespan, intervention planning must be both personalized and sensitive to developmental stage [15]. ADHD is not a static condition. Indeed, its clinical expression varies substantially between individuals and changes over time, influenced by developmental, environmental, and neurobiological factors [9, 14, 15]. Therefore, effective rehabilitation cannot rely on fixed protocols but must respond flexibly to each person's evolving cognitive and psychosocial needs. Although evidence-based cognitive rehabilitation programs have demonstrated their potential to improve executive functioning, particularly through structured exercises, behavioral compensatory strategies, and metacognitive techniques, their implementation in clinical practice remains challenging [17, 18].

First, clinicians must assess nuanced and often overlapping cognitive impairments, align their intervention with appropriate theoretical frameworks (*see, for instance*, the models of Barkley or Miyake & Friedman; 19, 20), and translate abstract concepts into practical, engaging, and sustainable formats. These demands are intensified by time constraints, limited resources (especially in some low-income countries), and a rather negative public attitude towards ADHD [18]. Specifically, there is currently a substantial imbalance between patients' needs and the limited availability of qualified specialists. This disparity has led to prolonged waiting periods, delayed diagnoses, and unequal access to effective treatment [18, 21]. As Ishikawa [22] points out, there is an urgent need to attract more young professionals into this field to meet the increasing demands for evaluation and management, particularly as awareness of the condition rises and diagnosis rates continue to grow. Finally, further complexity arises from the need to adapt interventions across the lifespan and comorbidities [10, 15, 17]. Indeed, what may be developmentally appropriate or motivating for a 15-year-old adolescent likely differs from what is feasible or relevant for a 65-year-old adult. Without age-sensitive adaptation, interventions risk being too simplistic or misaligned with the individual's context, reducing efficacy and engagement. A standardized, one-size-fits-all model thus fails to account for the nuanced cognitive, emotional, and contextual dynamics that influence rehabilitation outcomes at different life stages [9, 15].

The presence of overlapping symptoms, evolving diagnostic criteria, and variability in diagnostic tools and informant reports can further complicate the diagnostic process, despite the availability of effective and scalable tools for tailored interventions. Designing a personalized rehabilitation plan is, therefore, a resource-intensive task that often relies on the clinician's individual expertise and clinical intuition [17]. This limits the reproducibility of interventions and raises concerns about consistency and accessibility, particularly in under-resourced settings or among less experienced professionals.

The rise of AI in healthcare

Artificial intelligence (AI) has emerged as a transformative force in healthcare over the past few years. From diagnostics and decision-support systems to digital therapeutics and remote monitoring, AI-based tools are increasingly integrated into clinical workflows, promising to enhance efficiency, accuracy, and personalization (e.g., 23, 24, 25, 26). Within the domain of cognitive and mental health, AI has been applied to early detection of disorders, mood prediction, adaptive training platforms, and natural language processing for screening and intervention purposes [27, 28]. One of the most recent and high-profile developments is the emergence of generative AI, including large language models (LLMs) such as ChatGPT [29], which can generate human-like text in response to natural language prompts. These models have demonstrated versatility across various domains, including education, academic writing, customer service, and content creation [30].

In the context of personalized healthcare and the growing use of digital tools for remote support, the introduction of ChatGPT into neurorehabilitation settings presents a promising opportunity to address several existing challenges [31, 32]. By incorporating ChatGPT into therapeutic programs, clinicians can enhance traditional rehabilitation methods with AI-powered support. This integration may allow for a more adaptive and responsive form of care, where patients receive real-time, tailored guidance that can help sustain their motivation and active participation throughout the recovery journey [33]. ChatGPT, for example, can assist in tracking a patient's rehabilitation progress, suggesting appropriate exercises, and offering immediate, constructive feedback [32]. Moreover, because ChatGPT is easily accessible on digital devices such as smartphones and laptops, it provides a flexible and user-friendly platform that patients can use both within and outside of clinical settings [31, 33].

However, the use of these models for designing structured, clinically informed interventions, especially in neuropsychological rehabilitation, remains largely unexplored. This application introduces a set of significant and as-yet unresolved challenges. Key concerns include the accuracy of the generated content, its internal coherence, and its ability to align with individual cognitive profiles and developmental needs (e.g., 34, 35, 36, 37). Furthermore, interventions must adhere to established clinical and theoretical standards, which generative AI models may not consistently respect without appropriate guidance or validation. While these tools hold potential as scalable, flexible support for busy clinicians, particularly in contexts with limited time and resources, they also risk oversimplifying complex clinical phenomena or producing misleading outputs [32, 34–37]. Without rigorous evaluation, their integration into sensitive areas, such as cognitive rehabilitation, may compromise rather than enhance the quality of care. As such, the responsible exploration of generative AI's role in clinical program design requires critical scrutiny.

Study purpose and research questions

This study aims to investigate the potential of ChatGPT as a support system in the development of structured neuropsychological rehabilitation programs for individuals with ADHD. In particular, the study explores whether ChatGPT can generate interventions tailored to individual cognitive profiles and sensitive to the developmental differences that characterize ADHD across the lifespan. The focus is on enhancing executive functioning, a core area of impairment in ADHD, through personalized and theoretically grounded rehabilitation plans.

The following questions guide the research:

1. What experts in neurorehabilitation might ask ChatGPT to generate a rehabilitation protocol for the executive functions of individuals with ADHD?
2. How coherent, structured, and adaptable are the outputs generated by ChatGPT when applied to different ADHD profiles and developmental stages?
3. To what extent do clinical experts evaluate these AI-generated interventions as relevant, feasible, and consistent with established models of executive function rehabilitation?

By addressing these questions, the study seeks to clarify whether generative AI can offer meaningful support to clinicians in designing individualized intervention programs, serving not as a substitute for clinical expertise but as a complementary resource that enhances efficiency and personalization in therapeutic planning.

Study design overview

To explore the potential and limitations of AI-based support in neuropsychological rehabilitation planning for ADHD, the study was structured into three distinct yet interconnected studies. Each study addressed a specific objective, building upon the outcomes of the previous one: *Study 1*, a preliminary study focused on expert-driven prompt development; *Study 2*, which involved the generation and comparative analysis of AI-produced rehabilitation plans; and *Study 3*, which assessed the quality and applicability of these outputs through external expert evaluation. This tripartite structure allowed for an iterative approach to testing and refining AI-generated clinical content.

Specifically, in the first study, prompts were crafted using insights from clinical experts and existing literature (e.g., 34, 35, 36). These prompts were then assessed for quality, and the selected prompt was used to generate neuropsychological rehabilitation plans targeting executive functions with ChatGPT for three hypothetical ADHD profiles, each representing a different age group. The second study involved a semi-systematic comparison of the AI-generated plans, aimed at identifying commonalities, differences, and potential limitations across the profiles. This approach combines structured criteria with expert qualitative judgments, constituting a semi-systematic qualitative evaluation method that was named ISAAC, in honor of the mathematician Isaac Newton. This model represents a novel methodological contribution introduced in the present study, originally designed and developed by the authors. Finally, in the third study, a panel of clinical experts completed a semi-structured survey to assess the quality, relevance, and clinical applicability of the AI-generated content. The following sections provide a more detailed account of the methodology employed in each phase. The Institutional Ethical Board of the University of Palermo approved the study (N° 94232 – 2025).

Study 1 – Prompt generation based on expert input

Method

Participants

The first preliminary study focused on designing effective prompts for ChatGPT, informed by input from clinical experts. To this end, a total of 17 professionals with expertise in neuropsychology or neurorehabilitation (M = 5; F = 12; Mean age = 44.8; SD = 10.9) voluntarily participated in an online survey. Informed consent was obtained from all participants before their involvement, and the questionnaire was completed anonymously using Google Forms.

Materials and procedure

Data was collected through social media platforms and a snowball sampling method to reach a diverse group of experts. Following a brief set of sociodemographic questions, participants were asked to imagine planning an executive function rehabilitation program for a person with ADHD using ChatGPT. Specifically, they were presented with the following open-ended question: *"Imagine you have to plan an executive function rehabilitation intervention for a person with ADHD using ChatGPT. How would you structure your request (prompt)? Please write the exact sentence or sentences you would use."*

To further refine the prompt design process, experts were also asked to identify the type of information they would include in the prompt to personalize the AI-generated response in a clinically meaningful way. They were presented with a list of possible elements (e.g., age, severity of symptoms, daily life context) and asked to select up to three that they considered most important. Additional questions explored their expectations regarding the quality and content of a helpful AI-generated response. In this regard, experts were invited to select up to three key features they would expect from an effective output and up to three characteristics they believed a ChatGPT-generated rehabilitation program should have to be considered clinically useful (e.g., clarity, individualization, feasibility, etc.). The responses from this expert consultation were used to guide the construction of the final prompts submitted to ChatGPT in the next phase.

Data analysis

Socio-demographics questions and quantitative data from the closed-ended questions (e.g., selection of key information to include in prompts, expectations about AI-generated outputs) were analyzed using IBM SPSS Statistics (Version 26.0). Descriptive statistics were calculated to summarize the distribution of responses and highlight the most frequently selected options.

To analyze the open-ended responses collected in this first preliminary study, we adopted a structured mixed-method approach, drawing on the analytical framework proposed by Cortini and Tria [38]. According to their model, qualitative data can be explored through three complementary lenses: the narrative style adopted by respondents, the frequency and association of specific terms, and an integrated qualitative–quantitative perspective that captures both thematic and lexical patterns. For this study, we selected the third approach, as it enables a more comprehensive understanding of textual material by combining qualitative depth with quantitative structure, particularly useful when aiming to distill practical insights from expert-generated content. Following this approach, we first conducted an inductive thematic analysis of the responses to the open-ended question. Recurring themes and linguistic patterns were identified to extract shared strategies and expectations across experts' suggestions. To facilitate and support this process, we used NVivo [39], a qualitative data analysis software designed to help researchers organize, code, and analyze unstructured textual data. NVivo enabled the identification of word frequency patterns, clustering of similar responses, and the visualization of conceptual relationships. These analyses helped highlight which aspects of prompt design, such as the inclusion of specific executive functions, patient characteristics, or session structures, were most frequently prioritized by experts.

Prompt assessment and plan generation in Chat-GPT

To evaluate the quality of the prompts, a rapid assessment was conducted using three key criteria: Clarity (*Is the prompt understandable and unambiguous?);* Completeness (Does the prompt contain all essential elements, such as age, cognitive profile, and theoretical alignment?); Clinical relevance (*Is the prompt well-aligned to construct a practical cognitive rehabilitation intervention?*). Each prompt was then independently rated by the two authors on a 3-point scale (1 = low, 3 = high) for each of the three criteria. The cumulative score (maximum = 9) was used to identify the most effective prompt while minimizing response bias. This evaluation method enables a rapid and replicable comparison of prompt quality, without requiring AI-generated outputs. Given the absence of a standardized framework for clinical prompt evaluation, this approach was inspired by methods used in prompt engineering and educational technology research (e.g., 40, 41).

Following the independent ratings, the authors engaged in a discussion of their assessments. The ratings demonstrated a high degree of agreement, and no discrepancies arose that required resolution. The chosen prompt was subsequently adapted for other age groups, while maintaining its structure and clinical complexity unchanged. Specifically, it was submitted to ChatGPT-4o [29] for three different patient profiles (17-, 40-, and 65-year-old individuals with ADHD).

Results

Socio-demographics and quantitative data

Seventeen healthcare professionals, 29.4% males and 70.6% females, based in Italy, participated in the study, primarily from the regions of Sicily and Lombardy. The average age of participants was 44.8 years (SD = 10.9), with an average of over 15 years of professional experience in the field. Most reported working in private practice or rehabilitation centers and holding advanced degrees (postgraduate Master's or PhD) in psychology or a related field, such as neuroscience. Regarding AI familiarity, many participants indicated occasional or regular use of tools such as ChatGPT and reported a medium level of knowledge. Table 1 contains all relevant sociodemographic data.

Table 1 Sociodemographic characteristics of the sample (*N* = 17)

Variable	Category	Percemtage
Age	31–40 years	46.7%
	41–50 years	26.6%
	51–60 years	6.7%
	Over 60 years	20.0%
Gender	Male	29.4%
	Female	70.6%
Highest Educational Qualification	Postgraduate Master's Degree	53.3%
	PhD	20.0%
	Master of Science	13.3%
	MD	13.4%
Main Work Setting	Private Practice	66.7%
	Rehabilitation Center	13.3%
	Hospital/Clinical Center	13.3%
	University/Research Institute	6.7%
Years of Experience	5–10 years	40.0%
	11–15 years	13.3%
	16–20 years	20.0%
	Over 20 years	26.7%
Geographical Origin	Sicily	46.6%
	Lombardy	33.3%
	Piedmont	13.4%
	Sardinia	6.7%
Experience with ADHD Clients	Regular (5–10 cases/month)	46.7%
	Occasional (less than 5 cases/month)	40.0%
	Extensive (over 20 cases/month)	13.3%
AI Familiarity	Very High	6.7%
	High	20.0%
	Medium	33.3%
	Low	20.0%
	Poor	20.0%

In the closed-ended questions, three key elements emerged as the most frequently selected by participants when asked what information should be included in prompts to generate clinically useful AI responses. First, the patient's age and developmental stage were considered essential to ensure that the AI-generated output would be developmentally appropriate and contextually relevant. Second, experts emphasized that the rehabilitation program should be consistent with recognized theoretical models, reflecting established evidence-based practices. Third, they emphasized the need for personalization based on the patient's specific cognitive and functional profile, underscoring the importance of individualized interventions over generic content. These selections indicate a strong preference for clinically grounded, tailored approaches to AI-assisted rehabilitation. Further quantitative results, including frequency distributions and response breakdowns, are available in Supplementary File 1.

Thematic and content analysis regarding the prompt's generation question

The thematic analysis conducted using NVivo identified three main themes in participants' prompts to the AI system:

1. Planning Structured Rehabilitation Interventions: These prompts focused on designing structured cognitive rehabilitation programs targeting executive functions. Example: *"Plan a cognitive rehabilitation intervention for the executive functions of a 17-year-old boy diagnosed with ADHD."*
2. Practical Exercises and Activities: Participants also requested the generation of specific exercises, often emphasizing hands-on or operational components. Example: "*Create exercises for a patient with ADHD that stimulate executive functions, especially sequential mental planning."*
3. Guidelines and Theoretical Foundations: This theme included prompts referencing scientific standards or evidence-based practices. Example: *"What are the evidence-based guidelines for executive function rehabilitation in ADHD?"*

The content analysis examined 260 words of 17 answers, excluding articles and prepositions. The most frequently occurring terms were ADHD, with 12 occurrences (~ 4.6%), Executive Functions, with 12 occurrences (~ 4.6%), Rehabilitation, with 7 occurrences (~ 2.7%), and Intervention, with 5 occurrences (~ 1.9%). The full frequency table is available in Supplementary File 1.

Prompt generation and assessment

Based on the thematic analysis of both open- and closed-ended survey responses, three options of prompts were crafted to guide the development of clinically useful cognitive rehabilitation plans:

1) "Act as a cognitive rehabilitation professional. Design a personalized rehabilitation program that is consistent with recognized theoretical models for a 17-year-old patient with ADHD, taking into account their cognitive functioning level and any comorbid diagnoses. The program should include practical exercises and specific metacognitive strategies to enhance executive functions, with clear and progressive goals."
2) "Create rehabilitation exercises for a patient with ADHD aimed at improving executive functions, taking into account age and cognitive profile. The exercises should be clearly structured, applicable, and consistent with recognized theoretical models, including practical metacognitive strategies."
3) "Plan a cognitive rehabilitation intervention for a patient diagnosed with ADHD, focused on improving executive functions. Personalize the intervention based on the patient's age, cognitive profile, and specific goals. Ensure the program aligns with recognized theoretical models and includes practical activities and effective metacognitive strategies."

To identify the most effective input, three prompts were rated on clarity, completeness, and clinical relevance. The cumulative score (max = 9) guided prompt selection, independent of AI-generated outputs (Table 2).

Prompt 1 was therefore selected as the final AI input based on its highest overall score (9/9).

Discussion

This preliminary study provided a structured, transparent approach to designing and evaluating prompts for AI-assisted neuropsychological rehabilitation. By involving 17 anonymous clinical experts and combining qualitative and quantitative methods, the study ensured that the selected prompt reflected both clinical insight and methodological rigor. Unlike prior research (e.g., 34, 35, 36), which often lacks detail on how prompts or questions were generated, our process was explicitly documented and based on expert input, thematic analysis, and a mini framework for prompt assessment. For example, while Chen et al. [34] mention that questions were "crafted by a researcher and refined by a panel" (page 2), they do not report how these questions were selected or validated. Similarly, Latt et al. [36] derived inputs from telephone inquiries without detailing how they were summarized or rephrased by researchers, raising concerns about potential bias and reproducibility. In contrast, our three-step evaluation, which assessed clarity, completeness, and clinical relevance, provided a replicable method for selecting the most effective prompt. This approach responds to growing evidence that the quality and specificity of prompts can significantly impact the accuracy and utility of AI-generated outputs [40].

To conclude, by prioritizing transparency and clinical relevance, this phase laid the foundation for generating rehabilitation plans that are not only tailored but also methodologically sound. Yet, it is important to note that the prompts themselves were designed by the authors through several iterative steps, including literature review, expert consultation, and preliminary testing. Consequently, while our method provides a structured framework for prompt creation and assessment, the success of AI outputs relies heavily on the quality of the prompts. Therefore, clinicians must receive focused training in prompt creation, as this is essential to ensure that AI outputs are meaningful, accurate, and truly aligned with clinical goals.

Table 2 Assessment results of the prompts

Prompt	Clarity	Completeness	Clinical Relevance	Total
1 st Option	3	3	3	9
2nd Option	2	3	2	7
3rd Option	3	2	3	8

Study 2 – Comparison of AI-generated outputs

Method and procedure

After selecting the final prompt, the procedure for generating the AI-based rehabilitation plans was carried out in a controlled and replicable manner. Specifically, the prompt was submitted to ChatGPT-4o on two separate occasions (June 23, 2025, and July 7, 2025) by two different authors, each using a different laptop, IP address, and browser, without a logged-in account. This implies that each submission was performed on a different day and in a new chat session to reduce potential bias from session history or personalization effects. The prompt was used to generate plans for three distinct ADHD profiles (adolescent, adult, and older adult), and each prompt was entered twice, resulting in a total of six AI-generated rehabilitation plans.

To evaluate the quality of the generated outputs, a qualitative semi-systematic comparison framework developed explicitly for this study was employed. This approach combines structured evaluation criteria with the authors' judgment. While two key criteria were adapted from Xu et al. [35], the authors originally expanded and refined the framework by adding three additional criteria based on clinical reasoning and content analysis. This expansion represents a novel contribution to the framework, specifically tailored to the aims and context of the present study. The resulting framework, named *ISAAC*, includes five core dimensions that together allow for a structured comparison of AI-generated interventions:

- *Internal coherence* (e.g., logical consistency and progression; clarity of objectives, congruence between activities and objectives, explicit links to executive functions, and logical flow across phases);
- *Structural clarity* (e.g., presence of distinct and organized phases within the intervention; clear division into temporal phases, use of section titles, detailed scheduling of sessions, and clarity of language and transitions);
- *Accuracy* (e.g., alignment with current clinical knowledge; reference to theoretical models, use of validated psychometric tools, and avoidance of unvalidated or theoretically weak content);
- *Adaptability* (e.g., suitability for the targeted age group; relevance to adolescents, presence of personalization strategies, use of digital tools, and adaptation to ADHD-specific needs);
- *Comprehensiveness* (e.g., extent to which key intervention components are addressed; coverage of multiple executive functions, inclusion of metacognitive strategies and practical activities, integration of contextual factors like school or family, and presence of follow-up or generalization strategies).

The evaluation was initially conducted independently by the first author using this framework. A follow-up discussion was then held with the second author to compare impressions, resolve uncertainties, and refine the interpretation of scores. This collaborative process contributed to the reliability and consistency of the final assessments. All evaluations were recorded in a comparison table that allows for structured cross-analysis of the outputs.

Results

The six rehabilitation plans generated by ChatGPT (two responses each for three different ADHD profiles: adolescent, adult, and older adult) were evaluated using the ISAAC framework. For each profile, the second AI-generated output received the highest total score, as illustrated in Table 3.

Table 3 Summary of ISAAC evaluation scores by age profile

Age Profile	Response 1	Response 2	Highest Score
Adolescent	39	42	Response 2
Adult	36	38	Response 2
Older Adult	39	40	Response 2

While differences were often modest, the second output was consistently rated slightly higher in terms of internal coherence, language flow, and clarity. Notably, both responses across all profiles performed well in terms of Internal Coherence and Comprehensiveness, particularly in covering multiple executive functions and incorporating both practical and metacognitive strategies. However, Adaptability and Accuracy showed more variability, with lower scores on the inclusion of digital tools and explicit personalization.

Based on the overall scoring, Response 2 for all profiles was selected for expert review in the next phase (Study 3). The full evaluation tables are available in Supplementary File 2.

Discussion

The methodological precaution employed, consistent with best practices in the prompt engineering literature, was designed to minimize residual bias from session history or user-specific interactions (as suggested by Xu et al., 35), thereby contributing to the generation of high-quality outputs. Indeed, the results of Study 2 highlight the consistency and overall quality of ChatGPT's responses when guided by a well-structured and clinically grounded prompt. At the same time, the similarity in responses across different runs may reflect the high specificity of the prompt, which likely constrained variability in the model's output. This finding supports growing evidence (e.g., 40) that more specific and clinically detailed prompts tend to elicit more consistent and focused responses, reinforcing the critical role of prompt design in clinical applications of AI.

To systematically examine the quality and subtle distinctions among the outputs, the ISAAC framework was applied. While the differences across outputs were modest, the ISAAC framework proved essential in enabling a structured, clinically informed, and replicable comparison of the rehabilitation plans. Its systematic approach facilitated a nuanced evaluation that went beyond surface-level impressions, allowing for assessment of the internal coherence, progression structure, and therapeutic balance of each intervention. All plans demonstrated reasonable clinical plausibility, yet the highest-scoring responses distinguished themselves through well-organized phases and an effective integration of practical exercises with metacognitive strategies. Conversely, lower ratings in adaptability and contextual integration, such as limited personalization or insufficient reference to school or family environments, underscore areas where AI-generated interventions may fall short.

Study 3- external expert evaluation

Method

Participants and procedure

In the third study, a sample of 27 experts (M = 6; F = 21; mean age = 46.5; SD = 15) in neuropsychology and cognitive rehabilitation was recruited to independently evaluate the selected AI-generated intervention plan. Informed consent was obtained from all participants before their involvement, and the questionnaire was completed anonymously. The survey was created using Google Forms, and no email addresses or personal identifiers were collected during the process. The invitation was sent via email and contained a direct link to the survey. Distribution occurred through a mailing list provided by AIFA (Associazione Italiana Famiglie ADHD/Italian Association of ADHD Families), which included associations, healthcare professionals, and clinical centers across all Italian regions. More specifically, the mailing list targeted healthcare professionals involved in the cognitive rehabilitation of individuals of all ages with ADHD. This recruitment strategy ensured broad geographical coverage and diversity in clinical expertise.

Experts received the chosen plan for each age profile and completed a semi-structured survey assessing three core dimensions:

- *Clinical relevance* (e.g., usefulness, appropriateness, and potential applicability in real-world settings);
- *Feasibility* (e.g., ease of implementation considering typical constraints such as time and resources);
- *Theoretical consistency* (e.g., congruence with established cognitive models and rehabilitation frameworks for ADHD and executive functioning).

In addition to these dimensions, the survey also included basic demographic questions, items assessing participants' familiarity with and use of AI, as well as questions evaluating the potential of AI-generated content to serve as a helpful starting point for drafting neuropsychological rehabilitation plans. Furthermore, the survey explored the general strengths and limitations of AI outputs within the field of cognitive rehabilitation. Quantitative survey responses were analyzed using descriptive statistics in SPSS.

Results

Demographics

The sample comprises 27 Italian professionals working in the fields of psychology, psychiatry, and related clinical sciences. Most participants identified as female (77.8%), while 22.2% were male. The mean age was approximately 46 years (M = 46.5; SD = 15). Regarding educational background, 40.7% held a postgraduate master's degree, and 14.9% completed a PhD. Most respondents had an undergraduate education in Psychology or Cognitive Sciences (74.1%). Consistently, the most common areas of specialization were Clinical Psychology, Cognitive Sciences, Psychotherapy, and Neuropsychology.

The majority reported working primarily in private practice (70.1%), while 22.2% were employed in hospital or clinical settings, including community mental health services. A smaller portion (7.7%) worked in academic or educational institutions. In terms of professional experience, 25.9% had between 5 and 10 years of practice, and 25.9% had over 20 years of experience. Concerning clinical experience with clients diagnosed with ADHD, 51.9% reported working with such cases regularly (5–15 cases/month), while 29.6% reported an extensive level of engagement (> 15 cases/month). Finally, familiarity with AI varied across the sample. While 14.8% rated their familiarity at the lowest level, 59.3% rated themselves between 3 and 5 on a 5-point scale, indicating a moderate to high level of awareness. A complete summary of all demographic variables is presented in Table 4.

Evaluation of adolescent rehabilitation plan

Twenty-seven professionals evaluated the ChatGPT-generated intervention plan aimed at rehabilitating executive functions in a 17-year-old adolescent with ADHD. Their assessments focused on six key dimensions: clinical relevance, alignment with ADHD-related executive function challenges, consistency with current rehabilitation models, theoretical or empirical grounding of the proposed activities, feasibility in real-world settings, and the realism of the time, resources, and skills required for implementation.

As shown in Table 5, the intervention was rated as moderately to highly effective across most conceptual dimensions. The mean score for clinical relevance was 3.15 (SD = 0.78), suggesting moderate alignment with the characteristics of the ADHD adolescent population. The plan was perceived as adequately addressing executive function deficits typical of ADHD (M = 3.61, SD = 0.69), and it was judged to be largely coherent with current models of executive function rehabilitation (M = 3.65, SD = 0.68). Notably, the highest average score was assigned to the theoretical and empirical foundation of the proposed activities (M = 3.73, SD = 0.77). Conversely, the experts expressed greater reservations about the feasibility of the intervention. Ratings for implementation in real-world clinical settings averaged 2.5 (SD = 0.90), and the perceived realism of the time, resources, and competencies required scored similarly low, with a mean of 2.5 (SD = 0.98).

Despite these logistical concerns, only two participants reported the presence of inaccuracies or the use of unvalidated techniques. When asked whether they would feel comfortable using or adapting the intervention in their

Table 4 Sociodemographic characteristics of participants, including age, gender, education, work setting, professional experience, and familiarity with AI

Variable	Categories	Percentage
Age group	Under 30 years	7.5%
	31–40 years	40.7%
	41–50 years	14.8%
	51–60 years	14.8%
	Over 60 years	22.2%
Biological gender	Female	77.8%
	Male	22.2%
Highest educational qualification	Master's Degree (5-year or equivalent)	33.3%
	Postgraduate Master's (1–2 years after a master's degree)	40.7%
	PhD	14.9%
	Medical Degree (MD)	11.1%
Field of undergraduate degree	Psychology or Cognitive Sciences	74.1%
	Neuroscience or Neuropsychology	7.4%
	Psychiatry/Neurology/Neuropsychiatry	18.5%
Main work setting	Private Practice	70.1%
	Hospital/Clinical Setting	22.2%
	Academic/Educational Institution	7.7%
Italian region of work	Apulia	3.7%
	Campania	7.4%
	Emilia-Romagna	18.5%
	Friuli Venezia Giulia	3.7%
	Lazio	14.9%
	Lombardy	18.5%
	Piedmont	11.1%
	Sicily	11.1%
	Trentino-South Tyrol	3.7%
	Veneto	7.4%
Years of experience in the field	Less than 5 years	22.2%
	5–10 years	25.9%
	11–15 years	14.8%
	16–20 years	11.2%
	More than 20 years	25.9%
Experience with ADHD patients	Occasional (< 5 cases/month)	18.5%
	Regular (5–15 cases/month)	51.9%
	Extensive (> 15 cases/month)	29.6%
Familiarity with AI	1 (Not familiar at all)	14.8%
	2	25.9%
	3	33.3%
	4	14.8%
	5 (Very familiar)	11.2%

clinical work, 37% of respondents answered "Maybe," and another 37% "Probably yes," while a smaller group expressed reluctance, selecting "Definitely not" (14.8%) or "Probably not" (11.1%), as summarized in Table 6.

Evaluation of adult rehabilitation plan

Professionals also evaluated the ChatGPT-generated intervention plan designed to support executive function rehabilitation in a 40-year-old adult with ADHD.

Table 5 Mean expert ratings across evaluation dimensions (*N* = 24)

Evaluation Dimension	Mean	SD
Clinical relevance for the ADHD adolescent profile	3.15	0.78
Addresses core executive function issues typical of ADHD	3.61	0.69
Consistency with current models of EF rehabilitation	3.65	0.68
Theoretical or validated foundation of proposed activities	3.73	0.77
Feasibility of implementation in real-world clinical settings	2.5	0.90
Realism of required time, resources, and competencies	2.5	0.98

Table 6 Expert responses on accuracy and intention to use the intervention

Qualitative Item	Response Category	*N*	%
Noticed inaccuracies or use of non-validated techniques	No	25	92.6%
	Yes	2	7.4%
Would use/adapt the intervention in clinical work	Definitely not	4	14.8%
	Probably not	3	11.1%
	Maybe	10	37%
	Probably yes	10	37%
	Definitely yes	0	0%

Table 7 Mean expert ratings across evaluation dimensions (*N* = 24**)**

Evaluation Dimension	Mean	SD
Clinical relevance for the ADHD adult profile	3.16	0.81
Addresses core executive function issues typical of ADHD	3.41	0.77
Consistency with current models of EF rehabilitation	3.33	0.81
Theoretical or validated foundation of proposed activities	3.5	0.65
Feasibility of implementation in real-world clinical settings	2.75	0.94
Realism of required time, resources, and competencies	2.5	0.97

The assessment followed the same structure used for the adolescent-focused version and the exact six core dimensions. As shown in Table 7, the intervention received moderate to positive evaluations on conceptual dimensions. The clinical relevance for the target population was rated with a mean of 3.16 (SD = 0.81). Experts judged that the plan adequately addresses the core executive dysfunctions of ADHD (M = 3.41, SD = 0.77) and is generally consistent with current rehabilitation models (M = 3.33, SD = 0.81). The highest average rating was for the theoretical or validated grounding of the proposed activities (M = 3.5, SD = 0.65), indicating moderate to strong alignment with existing literature and practices. However, feasibility aspects received more cautious evaluations. The mean score for implementation feasibility in real-world clinical contexts was 2.75 (SD = 0.94), and the realism of time, resource, and competence requirements was rated slightly lower at 2.5 (SD = 0.97).

As reported in Table 8, six respondents (22.2%) indicated they had noticed inaccuracies or the use of non-validated methods in the proposed intervention. When asked whether they would feel comfortable using or adapting the intervention in their professional work, 37% responded "Probably yes" and 3.7% "definitely yes". A smaller group expressed strong reluctance, with 7.4% selecting "Definitely not" and 22.2% "Probably not".

Table 8 Expert responses on accuracy and intention to use the intervention

Qualitative Item	Response Category	*N*	%
Noticed inaccuracies or use of non-validated techniques	No	21	77.8%
	Yes	6	22.2%
Would use/adapt the intervention in clinical work	Definitely not	2	7.4%
	Probably not	6	22.2%
	Maybe	8	29.6%
	Probably yes	10	37%
	Definitely yes	1	3.7%

Evaluation of older adult rehabilitation plan

Finally, the same professionals evaluated the ChatGPT-generated intervention aimed at rehabilitating executive functions of an older adult (age 65+) with ADHD. The evaluation framework mirrored that used for younger age groups. As summarized in Table 9, expert judgments suggest moderate confidence in the conceptual and theoretical foundations of the plan. The mean rating for clinical relevance was 2.88 (SD = 0.98), while the ability to address executive function issues received a mean rating of 3.26 (SD = 0.60). Experts were more cautious in evaluating consistency with current executive function rehabilitation models (M = 3.11, SD = 0.76) and the degree of theoretical or empirical grounding (M = 3.23, SD = 0.76). As with previous versions of the plan, feasibility aspects elicited lower scores. The feasibility of implementation in real-world settings was rated at 2.42 (SD = 0.90), and the realism of time, resource, and skill demands was scored similarly at 2.38 (SD = 0.98).

Table 9 Mean expert ratings across evaluation Dimensions – Older adult ADHD plan ($N = 24$)

Evaluation Dimension	Mean	SD
Clinical relevance for the 65 + ADHD population	2.88	0.98
Addresses core executive function issues typical of ADHD	3.26	0.60
Consistency with current models of EF rehabilitation	3.11	0.76
Theoretical or validated foundation of proposed activities	3.23	0.76
Feasibility of implementation in real-world clinical settings	2.42	0.90
Realism of required time, resources, and competencies	2.38	0.98

As shown in Tables 5 and 10 out of 27 professionals (18.5%) reported inaccuracies or the use of unvalidated techniques. When asked about their willingness to use or adapt the plan in clinical practice, responses varied. A combined 59.2% indicated some level of openness (18.5% "Probably yes", 7.4% "Definitely yes", and 33.3% "Maybe"), while 40.8% were reluctant, selecting either "Probably not" (29.7%) or "Definitely not" (11.1%).

Table 10 Expert responses on accuracy and intention to use the intervention ($N = 24$)

Qualitative Item	Response Category	*N*	%
Noticed inaccuracies or unvalidated techniques	No	22	81.5%
	Yes	5	18.5%
Would use/adapt the intervention in clinical work	Definitely not	3	11.1%
	Probably not	8	29.7%
	Maybe	9	33.3%
	Probably yes	5	18.5%
	Definitely yes	2	7.4%

General questions about AI outputs

At the end of the survey, professionals were invited to reflect on the potential of AI for supporting the development of cognitive rehabilitation programs. Participants could select up to three perceived benefits and three limitations of AI-generated content from a predefined list and were then asked whether they would consider using AI as a starting point for designing personalized rehabilitation plans. Among the perceived benefits, the most frequently mentioned was the usefulness of AI-generated content as a clinical starting point, reported by 79% of participants. This was followed by the speed of content generation (67%), and the clarity and structure of the material (58%) and A smaller proportion of professionals highlighted adherence to known theoretical principles (46%), the use of accessible and understandable language (33%), and the potential for personalization (29%).

Regarding limitations, the most frequently cited concern was the excessive generality of the proposed activities, noted by 83% of respondents. This was followed by concerns about limited real-world applicability (71%) and the absence of more explicit scientific references (42%). Additional limitations included the risk of inaccurate or fabricated content (38%), superficial theoretical depth (33%), insufficient adaptation to specific clinical conditions (25%), and the use of language that was either too technical or overly simplistic (25%).

Overall, despite the identified limitations, the responses revealed a cautiously open attitude toward integrating AI tools in clinical practice. Specifically, 7.4% of participants stated they would definitely consider using AI as a support tool, while 33.3% responded *probably yes*, and 29.6% indicated *maybe*. Meanwhile, 22.2% expressed some reluctance, answering *"probably not,"* and a smaller proportion (7.4%) stated that they would *not* consider using such tools (Fig. 1).

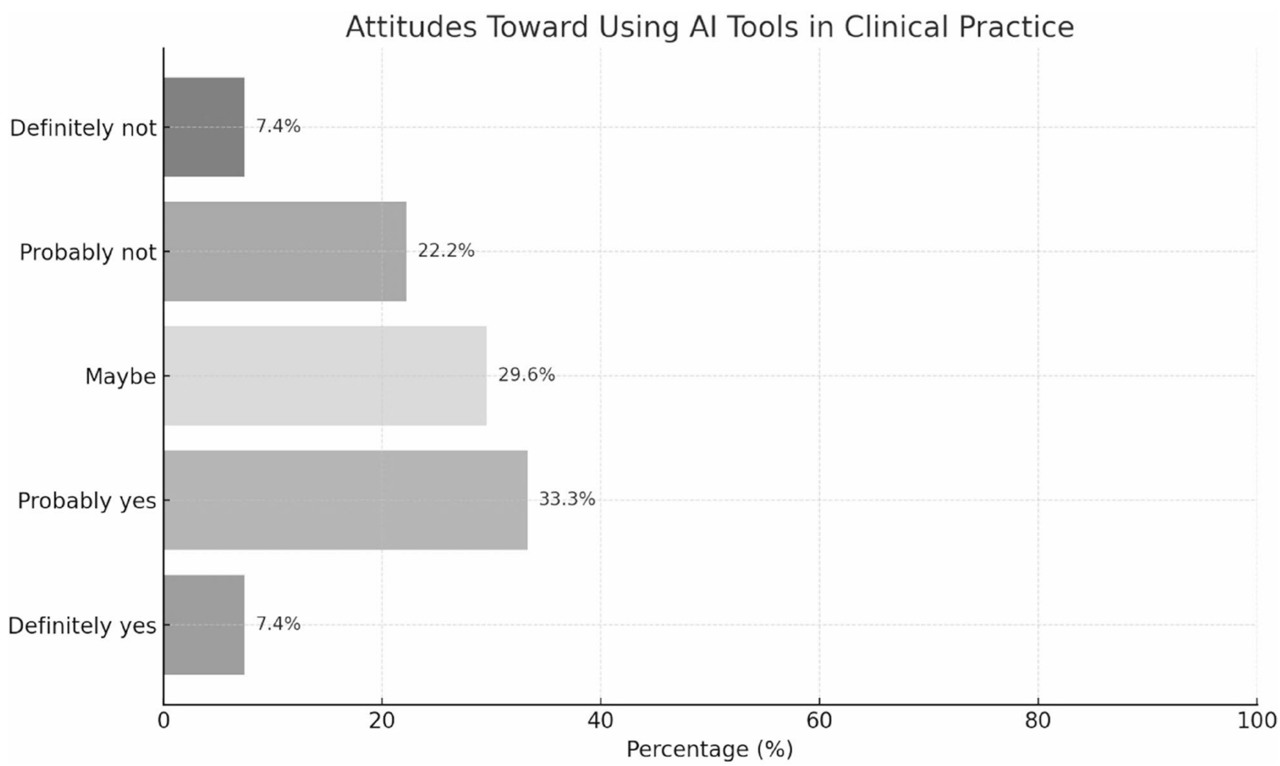

Fig. 1 Participants' general attitudes toward the use of AI in clinical practice as a support tool at the end of the survey

Discussion

The evaluations indicate a generally positive reception of the AI-generated intervention plans across all age groups, particularly in terms of their conceptual validity and alignment with the executive function challenges associated with ADHD. Experts recognized the theoretical coherence of the interventions, especially for adolescent and adult profiles. Furthermore, the overall evaluation experience was positive. Indeed, many professionals stated they would consider using AI tools like ChatGPT as a supportive aid in planning cognitive rehabilitation programs, reflecting a cautiously optimistic attitude toward AI integration in clinical practice. A minority of professionals, however, noted instances of inaccuracies or the inclusion of unvalidated techniques, particularly in plans targeting adult and older adult populations. According to the authors' view, this may be partly explained by the fact that ADHD is still largely conceptualized as a developmental disorder, with a stronger evidence base and clinical focus on childhood and adolescence. As a result, there is comparatively less structured guidance for adult and older adult populations, which may have contributed both to the limitations in the AI-generated outputs and to experts' more critical evaluations of these profiles.

Moreover, feasibility remains the primary concern. Across all versions, implementation in real-world settings, personalization, and the realism of required resources received lower ratings. These findings echo earlier results from other healthcare domains. For instance, when ChatGPT was tasked with generating occupational therapy plans based on clinical case vignettes, experts acknowledged the overall structure and coherence of the outputs. However, they raised concerns about their lack of personalization and insufficient clinical detail [42]. Similarly, broader evaluations of ChatGPT in clinical decision-making contexts have highlighted that, while the content may appear plausible, it often lacks the nuance, accuracy, and feasibility necessary for reliable real-world use, leading to cautious skepticism among healthcare providers [43].

Finally, we observed that, while experts primarily focused on personalization and accuracy, data security and confidentiality were considered less critical in this study. All prompts and outputs were fully anonymized, and no patient-identifiable information was used. Nevertheless, in real-world clinical applications, attention to privacy, data protection, and secure storage of patient information would be essential.

General discussion

Although the use of AI in psychiatric and neurological fields is receiving increasing recognition (e.g., 23, 24, 25, 26, 44), its application within neurorehabilitation remains largely underexplored [32], particularly regarding ADHD. This study thus explored whether generative AI, specifically ChatGPT-4o, can meaningfully support clinicians in designing individualized rehabilitation plans for executive function deficits in individuals with ADHD. It focused on identifying a specific prompt to generate such interventions, evaluating, then, the coherence, structure, and adaptability of the resulting outputs across various ADHD profiles and developmental stages. The outputs produced by the model were assessed using ISAAC, a novel framework explicitly developed by the authors for this study to evaluate the quality of AI-generated rehabilitation content systematically. ISAAC includes five key dimensions: internal coherence, structural clarity, accuracy, adaptability, and comprehensiveness. Based on this framework, the most coherent and well-structured responses were selected for external expert evaluation. A panel of clinicians subsequently reviewed these selected plans, assessing their relevance, feasibility, and alignment with established models of executive function rehabilitation.

Findings showed that ChatGPT, when guided by carefully designed prompts, can generate neuropsychological rehabilitation plans that are conceptually valid and theoretically grounded. Experts consistently rated the plans positively in terms of addressing core executive function challenges and aligning with current rehabilitation practices. Nevertheless, feasibility emerged as a significant and recurrent concern. Despite the conceptual soundness and clinical relevance of the AI-generated interventions, experts often questioned whether these plans could realistically be implemented with the intended populations. These concerns emerged particularly in the open comment section at the end of the survey, where clinicians were invited to share additional reflections. For example, one clinician remarked: "*It is difficult to change the chronic habits of a 65-year-old ADHD patient*." While another commented, "*ADHD patients are not able to follow through with a program, so I find it difficult to imagine planning out a week using a fixed schedule. They get bored easily*."

These comments reflect a broader tension between theoretical planning and the realities of clinical engagement. Even when AI-generated plans are coherent and aligned with best practices, they may require considerable professional adaptation to accommodate motivational challenges, cognitive variability, and resource limitations common in ADHD populations. As such, the findings reaffirm that AI should be viewed as a supportive resource, capable of accelerating idea generation and aiding decision-making, but reliant on expert oversight to ensure ethical, personalized, and context-sensitive implementation [32, 45].

Clinical and ethical considerations

In addition to feasibility concerns, the integration of generative AI into cognitive rehabilitation raises several broader clinical and ethical issues that must be carefully addressed. One key clinical implication is the risk of deskilling and overreliance [46, 47]. As AI systems become increasingly capable of producing structured and clinically plausible interventions, there is a risk that clinicians, particularly those in training, at the beginning of their careers, or under pressure, may begin to rely on these tools without critically evaluating or adapting their outputs. This overreliance could also undermine not only the development but also the training of essential clinical competencies, as the brain, like muscle, requires regular exercise to strengthen skills such as individualized formulation, flexible treatment planning, and therapeutic judgment [46].

Another concern is the issue of contextual and cultural insensitivity. LLMs generate responses based on generalized patterns learned from vast and often culturally homogeneous datasets. As a result, they may fail to account for nuances related to socio-cultural background, language, family dynamics, and educational or occupational settings, factors that are central to neuropsychological rehabilitation [48]. Without careful human mediation, AI-generated interventions may unintentionally replicate existing biases or overlook the needs of underrepresented populations, thereby exacerbating healthcare disparities [49, 50]. Additionally, AI systems often lack the ethical reasoning and emotional intelligence that are essential to clinical care [51, 52]. Decisions involving risk management, patient autonomy, informed consent, and the therapeutic alliance also require a nuanced understanding of the patient's emotional state, personal values, and relational context, domains where AI has no genuine competence [49, 51, 52]. For instance, deciding whether to involve family members in a rehabilitation plan or how to address a patient's ambivalence toward treatment in a sensitive manner demands human empathy, discretion, and moral judgment [51]. These complex, context-dependent decisions cannot be reliably replicated by AI, underscoring that such systems can only complement, not replace, expert clinical judgment [32, 48, 53].

Ultimately, there is a need to consider the evolving relationship between clinicians and AI. As these tools become more integrated into practice, it will be essential to define boundaries, roles, and responsibilities [54–56]. Clear policies must distinguish between AI-assisted decision-making and autonomous clinical judgment [55, 57]. For example, institutional guidelines might require clinicians to document any AI-generated recommendations they used explicitly and to provide a rationale for their acceptance or modification, ensuring that ultimate responsibility remains with the human professional. Without such frameworks, there is a risk of ethical ambiguity or role confusion, particularly when adverse outcomes occur or treatment plans prove ineffective [57]. This is particularly important because poor decisions in healthcare, whether made by clinicians or AI systems, can lead to a breakdown of patient trust, widen existing healthcare disparities, and ultimately result in harmful outcomes for individuals [58, 59].

Future directions

The integration of generative AI into neuropsychological rehabilitation presents both exciting opportunities and critical challenges. Looking ahead, one of the most promising applications of these tools lies in early-stage clinical planning, as demonstrated in this study. However, realizing their full potential will require a targeted research agenda. Future studies should continue evaluating the clinical effectiveness of AI-generated interventions in real-world settings. This includes controlled trials aimed at comparing AI-supported plans to traditional rehabilitation approaches (designed exclusively by human professionals) to assess specific strengths and weaknesses, as well as longitudinal studies assessing outcomes such as adherence, functional improvement, and patient satisfaction. Additionally, studies could examine the perspectives of end-users and patients regarding the usefulness, accessibility, and motivational impact of AI-generated plans. Understanding patient and clinician experiences will help optimize AI integration and ensure interventions are practical, acceptable, and engaging.

Ultimately, research should investigate the optimal structure for human-AI collaboration in clinical settings. Questions remain about when and how clinicians should rely on AI input, how to maintain professional responsibility, and how to foster trust without overdependence. Developing evidence-based frameworks to support ethical and practical use of generative AI in rehabilitation will be essential as these tools become more embedded in mental health practice. For this reason, there is also a need to investigate how different user profiles, such as students, early-career clinicians, and experienced professionals, interact with AI tools and integrate them into their decision-making processes.

Limits and strengths of the study

This study represents the first systematic attempt to explore the potential of AI in generating neuropsychological rehabilitation plans tailored to individuals of different ages with ADHD, focusing on executive functions. A key strength lies in the multi-methodological approach, which combines prompt generation, analysis, and selection of the most effective prompt, AI content generation, structured evaluation through the newly developed ISAAC framework, and a final assessment by practicing clinicians. The design paid particular attention to minimizing bias in both the prompt generation phase, through expert consultation and prompt quality assessment, and the output evaluation phase, using independent scoring and controlled AI input procedures. This methodological transparency and triangulation across phases contribute to the study's reliability and replicability.

Nonetheless, some limitations should be acknowledged. The sample size was relatively small and may not capture the full range of professional perspectives. Additionally, while experts were drawn from across Italy, the majority of participants in the third study were psychologists, resulting in an underrepresentation of other key professional groups such as child neuropsychiatrists, speech-language pathologists, occupational therapists, and psychiatric rehabilitation technicians, who might be involved in ADHD care. In addition, although the evaluation of AI-generated outputs was carefully structured, including a third-party external evaluator might have enhanced objectivity and strengthened the consensus on the best output. Finally, although clinical experts rigorously evaluated AI-generated outputs, no pilot studies or trials involving real patients were conducted. Consequently, the study cannot determine how these AI-generated rehabilitation plans perform in practical clinical settings. This limitation affects the ability to assess critical outcomes such as patient adherence, engagement, functional improvements, and overall clinical impact.

Conclusions

This study offers preliminary insights into the potential of AI to support the design of clinically relevant and theoretically coherent neuropsychological rehabilitation plans for individuals with ADHD across different age groups. Methodologically, it highlights the value of structured prompt design and systematic expert evaluation in ensuring the clinical soundness of AI-generated materials. Prompt formulation should be grounded in established therapeutic reasoning, and expert review remains essential to assess both the quality and applicability of AI outputs. At the same time, the findings emphasize the need for cautious interpretation. Although experts expressed interest in the potential use of AI-assisted tools, concerns regarding feasibility, validation, and integration into existing clinical workflows remain substantial.

In conclusion, AI-assisted rehabilitation design appears to be a promising yet still exploratory field. Future work should focus on empirical validation, interdisciplinary collaboration, and ethical oversight to ensure that such technologies complement, rather than replace, human clinical expertise in mental health care.

Abbreviations

ADHD Attention–deficit/hyperactivity disorder
AI Artificial Intelligence
ISAAC Internal coherence, structural clarity, accuracy, adaptability, and comprehensiveness (The qualitative evaluation framework developed by the authors used to assess AI–generated intervention plans)
LMM Large multimodal model

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40359-025-03729-2.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The authors wish to thank AIFA (Associazione Italiana Famiglie ADHD – Italian Association of ADHD Families) for their valuable support in providing access to a mailing list of professionals specialized in cognitive and emotional rehabilitation for individuals with ADHD.

Authors' contributions

MD : conceptualization; methodology; data collection; data analysis (qualitative and quantitative); writing – original draft; writing – review and editing; submission. **BC** : conceptualization; methodology; data collection; supervision; writing – review and editing.

Funding

No external funding was received for this study.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to confidentiality agreements and the sensitive nature of participant data, but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The Institutional Ethical Board of the University of Palermo approved the study (N° 94232 – 2025). All participants provided written informed consent prior to participation.

Competing interests

The authors declare no competing interests.

Consent for publication

All participants provided consent for the publication of anonymized data as part of the informed consent process.

Received: 1 August 2025 / Accepted: 13 November 2025
Published online: 05 December 2025

References

1. Quintero-López C, Gil-Vera VD, Landinez-Martínez DA, Vargas Gaviria JP, Gómez-Muñoz N. Predictive neurocognitive model of attention deficit hyperactivity disorder diagnosis. Mediterr J Clin Psychol. 2023;11(1). https://doi.org/10.13129/2282-1619/mjcp-3606.
2. Abdelnour E, Jansen MO, Gold JA. ADHD diagnostic trends: increased recognition or overdiagnosis? Mo Med. 2022;119(5):467–73.
3. Salari N, Ghasemi H, Abdoli N, Rahmani A, Shiri MH, Hashemian AH, Akbari H, Mohammadi M. The global prevalence of ADHD in children and adolescents: a systematic review and meta-analysis. Ital J Pediatr. 2023;49(1):48. https://doi.org/10.1186/s13052-023-01456-1.
4. Song P, Zha M, Yang Q, Zhang Y, Li X, Rudan I, et al. The prevalence of adult attention-deficit hyperactivity disorder: a global systematic review and meta-analysis. J Glob Health. 2021;11:04009. https://doi.org/10.7189/jogh.11.04009.
5. Martino G, Caprì T, Castriciano C, Fabio RA. Automatic deficits can lead to executive deficits in ADHD. Mediterr J Clin Psychol. 2017;5(3). https://doi.org/10.6092/2282-1619/2017.5.1669.
6. Mohamed SMH, Butzbach M, Fuermaier ABM, Weisbrod M, Aschenbrenner S, Tucha L, et al. Basic and complex cognitive functions in adult ADHD. PLoS One. 2021;16(9):e0256228. https://doi.org/10.1371/journal.pone.0256228.
7. Cristofori I, Cohen-Zimerman S, Grafman J. Executive functions. Handb Clin Neurol. 2019;163:197–219. https://doi.org/10.1016/B978-0-12-804281-6.00011-2.
8. Jong A, Odoi CM, Lau J, Hollocks MJ. Loneliness in young people with ADHD: a systematic review and meta-analysis. J Atten Disord. 2024;28(7):1063–81. https://doi.org/10.1177/10870547241229096.
9. Leffa DT, Caye A, Rohde LA. ADHD in children and adults: diagnosis and prognosis. In: Stanford SC, Sciberras E, editors. New discoveries in the behavioral neuroscience of attention-deficit hyperactivity disorder. Curr top behav neurosci. Volume 57. Cham: Springer; 2022. https://doi.org/10.1007/7854_2022_329.
10. Fischer S, Nilsen C. ADHD in older adults – a scoping review. Aging Ment Health. 2024;28(9):1189–96. https://doi.org/10.1080/13607863.2024.2339994.
11. Thorell LB, Holst Y, Sjöwall D. Quality of life in older adults with ADHD: links to ADHD symptom levels and executive functioning deficits. Nord J Psychiatry. 2019;73(7):409–16. https://doi.org/10.1080/08039488.2019.1646804.
12. Quintero-López C, Gil-Vera VG, Ochoa JV, Herrera Velez A. The effect of working memory on social skills in male children with ADHD: A structural equation model. Mediterr J Clin Psychol. 2024;12(2). https://doi.org/10.13129/2282-1619/mjcp-4130.
13. Cherkasova M, Sulla EM, Dalena KL, Pondé MP, Hechtman L. Developmental course of attention deficit hyperactivity disorder and its predictors. J Can Acad Child Adolesc Psychiatry. 2013;22(1):47–54.
14. Di Lorenzo R, Balducci J, Poppi C, Arcolin E, Cutino A, Ferri P, et al. Children and adolescents with ADHD followed up to adulthood: a systematic review of long-term outcomes. Acta Neuropsychiatr. 2021;33(6):283–98. https://doi.org/10.1017/neu.2021.23.
15. Leahy LG. Diagnosis and treatment of ADHD in children vs adults: what nurses should know. Arch Psychiatr Nurs. 2018;32(6):890–5. https://doi.org/10.1016/j.apnu.2018.06.013.
16. Leffa DT, Ferrari-Souza JP, Bellaver B, et al. Genetic risk for attention-deficit/hyperactivity disorder predicts cognitive decline and development of Alzheimer's disease pathophysiology in cognitively unimpaired older adults. Mol Psychiatry. 2023;28:1248–55. https://doi.org/10.1038/s41380-022-01867-2.
17. Koyuncu A, Ayan T, İnce Guliyev E, Erbilgin S, Deveci E. ADHD and anxiety disorder comorbidity in children and adults: diagnostic and therapeutic challenges. Curr Psychiatry Rep. 2022;24(2):129–40. https://doi.org/10.1007/s11920-022-01324-5.
18. Zysset A, Robin D, Albermann K, Dratva J, Hotz S, Wieber F, von Rhein M. Diagnosis and management of ADHD: a pediatric perspective on practice and challenges in Switzerland. BMC Pediatr. 2023;23(1):103. https://doi.org/10.1186/s12887-023-03873-x.
19. Barkley RA. A theory of ADHD. In: Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment. 2006. pp. 297–334; https://doi.org/10.1177/1087054707305334
20. Hoadley R, Corley E, Russ KF, Pastuszak JP, Hill BD. Executive functioning in neurologic disorders: applying the Miyake and Friedman model to medical patients. In: Hill BD, editor. Executive functioning. New York: Springer; 2014. pp. 49–82.
21. Coppini V, Ferraris G, Ferrari MV, Dahò M, Kirac I, Renko I, et al. Patients' perspectives on cancer care disparities in Central and Eastern European countries: experiencing taboos, misinformation, and barriers in the healthcare system. Front Oncol. 2024;14:1420178. https://doi.org/10.3389/fonc.2024.1420178.
22. Ishikawa M. Changes in the characteristics of rehabilitation physicians over two decades: analysis of national physician census surveys in Japan. Prog Rehabil Med. 2020;5:20200012. https://doi.org/10.2490/prm.20200012.
23. Bohr A, Memarzadeh K, editors. Artificial intelligence in healthcare. Academic. 2020. https://shop.elsevier.com/books/artificial-intelligence-in-healthcare/bohr/978-0-12-818438-7.

24. Cheng SW, Chang CW, Chang WJ, Wang HW, Liang CS, Kishimoto T, et al. The now and future of ChatGPT and GPT in psychiatry. Psychiatry Clin Neurosci. 2023;77:592–6. https://doi.org/10.1111/pcn.13588.
25. Coppini V, Ferraris G, Ferrari MV, Monzani D, Dahò M, Fragale E, et al. The Beacon Wiki: mapping oncological information across the European Union. BMC Med Inform Decis Mak. 2025;25(1):193. https://doi.org/10.1186/s12911-025-03015-6.
26. Saraswat D, Bhattacharya P, Verma A, Prasad VK, Tanwar S, Sharma G, et al. Explainable AI for healthcare 5.0: opportunities and challenges. IEEE Access. 2022;10:84486–517. https://doi.org/10.1109/ACCESS.2022.3197671.
27. Baran FDE, Cetin M. AI-driven early diagnosis of specific mental disorders: a comprehensive study. Cogn Neurodyn. 2025;19(1):70. https://doi.org/10.1007/s11571-025-10253-x.
28. Sun J, Dong QX, Wang SW, Zheng YB, Liu XX, Lu TS, et al. Artificial intelligence in psychiatry research, diagnosis, and therapy. Asian J Psychiatr. 2023;87:103705. https://doi.org/10.1016/j.ajp.2023.103705.
29. OpenAI. ChatGPT [Large language model]. 2023. Available from: https://chat.openai.com/chat
30. Akhtar ZB. Unveiling the evolution of generative AI (GAI): a comprehensive and investigative analysis toward LLM models (2021–2024) and beyond. J Electr Syst Inf Technol. 2024;11:22. https://doi.org/10.1186/s43067-024-00145-1.
31. Eriksen AV, Soren M, Jesper R. Use of GPT-4 to diagnose complex clinical cases. NEJM. 2024;1. https://doi.org/10.1056/AIp2300031.
32. Maggio MG, Tartarisco G, Cardile D, Bonanno M, Bruschetta R, Pignolo L, et al. Exploring chatgpt's potential in the clinical stream of neurorehabilitation. Front Artif Intell. 2024;7:1407905. https://doi.org/10.3389/frai.2024.1407905.
33. Chakraborty C, Pal S, Bhattacharya M, Dash S, Lee SS. Overview of chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. Front Artif Intell. 2023;6:1237704. https://doi.org/10.3389/frai.2023.1237704.
34. Chen YC, Lee SH, Sheu H, Lin SH, Hu CC, Fu SC, et al. Enhancing responses from large language models with role-playing prompts: a comparative study on answering frequently asked questions about total knee arthroplasty. BMC Med Inform Decis Mak. 2025;25:196. https://doi.org/10.1186/s12911-025-03024-5.
35. Xu Q, Wang J, Chen X, Wang J, Li H, Wang Z, Li W, Gao J, Chen C, Gao Y. Assessing the efficacy of ChatGPT prompting strategies in enhancing thyroid cancer patient education: a prospective study. J Med Syst. 2025;49(1):11. https://doi.org/10.1007/s10916-024-02129-0.
36. Latt PM, Aung ET, Htaik K, Soe NN, Lee D, King AJ, et al. Evaluation of artificial intelligence (AI) chatbots for providing sexual health information: a consensus study using real-world clinical queries. BMC Public Health. 2025;25(1):1788. https://doi.org/10.1186/s12889-025-22933-8.
37. Tangsrivimol JA, Darzidehkalani E, Virk HUH, Wang Z, Egger J, Wang M, et al. Benefits, limits, and risks of ChatGPT in medicine. Front Artif Intell. 2025;8:1518049. https://doi.org/10.3389/frai.2025.151804.
38. Cortini M, Tria S. Triangulating qualitative and quantitative approaches for the analysis of textual materials: an introduction to T-lab. Soc Sci Comput Rev. 2014;32(4):561–8. https://doi.org/10.1177/08944393135101.
39. Jackson K, Bazeley P. Qualitative data analysis with NVivo. 3rd ed. London: Sage; 2019.
40. Lee D, Palmer E. Prompt engineering in higher education: a systematic review to help inform curricula. Int J Educ Technol High Educ. 2025;22(1):7. https://doi.org/10.1186/s41239-025-00503-7.
41. Reynolds L, McDonell K. Prompt programming for large language models: Beyond the few-shot paradigm. In: Extended abstracts of the 2021 CHI conference on human factors in computing systems. 2021; pp. 1–7; https://doi.org/10.48550/arXiv.2102.07350
42. Inoue T, Sawamura S, Sera T, Takenaka T, Kohiyama K, Nagai T. ChatGPT and occupational therapy: a study of generated program feasibility. Cureus. 2025;17(5):e83761. https://doi.org/10.7759/cureus.83761.
43. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst. 2023;47(1):33. https://doi.org/10.1007/s10916-023-01925-4.
44. Dahò M, Monzani D. The multifaceted nature of inner speech: phenomenology, neural correlates, and implications for aphasia and psychopathology. Cogn Neuropsychol. 2025. https://doi.org/10.1080/02643294.2025.2527983.
45. Mittal K, Dhar M. Use of ChatGPT by physicians to build rehabilitation plans for the elderly: a mini-review of case studies. J Indian Acad Geriatr. 2023;19(2):86–93. https://doi.org/10.4103/0974-3405.380506.
46. Natali C, Marconi L, Dias Duran LD, Miglioretti M, Cabitza F. AI-induced deskilling in medicine: a mixed method literature review for setting a new research agenda. SSRN [Preprint]. 2025. Available from: https://ssrn.com/abstract=5166364
47. Shukla P, Bui P, Levy SS, Kowalski M, Baigelenov A, Parsons P. De-skilling, cognitive offloading, and misplaced responsibilities: potential ironies of AI-assisted design. In: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 2025; Art. N. 171, Pg. 1–7; https://doi.org/10.1145/3706599.371993
48. Davies GK, Davies MLK, Adewusi E, Moneke K, Adeleke O, Mosaku LA, et al. AI-enhanced culturally sensitive public health messaging: a scoping review. E-Health Telecommun Syst Netw. 2024;13(4):45–66. https://doi.org/10.4236/etsn.2024.134004.
49. Amaro V, Zhang X. Intercultural interfaces: artificial intelligence and the challenges of cultural sensitivity. E-Revista De Estudos Interculturais. 2025;(13). https://www.iscap.pt/cei/e-rei/e-rei/n13/artigo18.pdf.
50. Gupta M, Parra CM, Dennehy D. Questioning racial and gender bias in AI-based recommendations: do espoused National cultural values matter? Inf Syst Front. 2022;24(5):1465–81. https://doi.org/10.1007/s10796-021-10156-2.
51. da Silva TMHR. Emotional intelligence and AI in geriatric nursing: bridging technology and compassionate care. In: Physical health, mental health, and human well-being in the age of AI. Hershey, PA: IGI Global; 2024. p. 199–232. https://doi.org/10.4018/979-8-3693-6190-0.ch010.
52. Ghotbi N. The ethics of emotional artificial intelligence: a mixed method analysis. Asian Bioeth Rev. 2022;15(4):417–30. https://doi.org/10.1007/s41649-022-00237-y.
53. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018;28:689–707. https://doi.org/10.1007/s11023-018-9482-5.
54. Barclay I, Abramson W. Identifying Roles, Requirements, and responsibilities in trustworthy AI systems. ArXiv [Preprint]. 2021. https://doi.org/10.48550/arXiv.2106.08258.
55. Giansanti D. The regulation of artificial intelligence in digital radiology in the scientific literature: a narrative review of reviews. Healthcare. 2022;10(10):1824. https://doi.org/10.3390/healthcare10101824.
56. Westerstrand S. Reconstructing AI ethics principles: Rawlsian ethics of artificial intelligence. Sci Eng Ethics. 2024;30(5):1–21. https://doi.org/10.1007/s11948-024-00507-y.
57. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. Heliyon. 2024. https://doi.org/10.1016/j.heliyon.2024.e26297.
58. Dahò M. Emotional responses in clinical ethics consultation decision-making: an exploratory study. Behav Sci. 2025;15(6):748. https://doi.org/10.3390/bs15060748.
59. Dahò M, Coppini V, Ferrari MV, Ferraris G, Sanchini V, Monzani D, et al. Exploring barriers to inclusivity: systematic analysis of exclusion criteria and potential bias in clinical cancer trials for psychiatric and neurological conditions in European protocols. Psychooncology. 2025;34(5):e70182. https://doi.org/10.1002/pon.70182.

Publisher's Note