MUNI
FSS

JOURNAL OF PSYCHOSOCIAL RESEARCH ON CYBERSPACE
CYBERPSYCHOLOGY

Buattini, M., Barjami, D., Paponetti, L., Torres, D., Borlimi, R., & Caselli, G. (2026). Evaluating strengths, limitations, and future directions of ChatGPT in psychological analysis within case conceptualization: A qualitative analysis. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 20*(1), Article 4.
https://doi.org/10.5817/CP2026-1-4

Evaluating Strengths, Limitations, and Future Directions of ChatGPT in Psychological Analysis Within Case Conceptualization: A Qualitative Analysis

Matilde Buattini[1,2,3], *Donald Barjami*[1], *Lorenza Paponetti*[1], *Dalila Torres*[1,2], *Rosita Borlimi*[4], *& Gabriele Caselli*[1,2,3]

[1] MeThe Research Lab, Psychology Department, Sigmund Freud University Milano, Milan, Italy
[2] Psychology Department, Sigmund Freud Privat Universität, Wien, Austria
[3] Studi Cognitivi, Milan, Italy
[4] Affective Neuroscience Lab, Psychology Department, Sigmund Freud University Milano, Milan, Italy

Abstract

This exploratory qualitative study investigates ChatGPT-4's capacity to apply the LIBET case formulation model by analyzing its feedback on anonymized interview transcripts. The study aimed to assess whether ChatGPT-4's outputs reflected accurate identification and interpretation of two key psychological constructs—life themes and semi-adaptive plans—while adhering to theoretical principles, and to explore recurring errors and limitations in its clinical reasoning. Ten non-clinical participants underwent semi-structured interviews, and a custom-configured version of ChatGPT-4 was provided with structured instructions and theoretical material. Reflexive thematic analysis revealed four overarching themes: (1) limitations in abstraction and interpretative barriers, (2) consistent structure and content organization, (3) hypothesis-driven reasoning with cautious language, and (4) partial adherence to LIBET theory through appropriate terminology. While ChatGPT's structured reasoning and alignment with theoretical vocabulary suggest its potential as a reflective support tool—particularly in training or supervision—it also showed difficulties in distinguishing emotional vulnerabilities from coping strategies, and in interpreting abstract, relational constructs such as life themes. Findings support the importance of improving prompt design, expanding training on psychological constructs, and developing rigorous validation pipelines. Future research should address these limitations before deploying LLMs as assistive tools in clinical reasoning and decision-making.

Keywords: ChatGPT; case formulation; case conceptualization; large language models; thematic analysis

Editorial Record

First submission received:
January 29, 2025

Revisions received:
June 19, 2025
November 12, 2025

Accepted for publication:
November 26, 2025

Editor in charge:
Lenka Dedkova

Introduction

Natural Language Processing (NLP) consists of algorithms capable of understanding, analyzing, and representing natural language. The use of NLP in psychotherapy has shown promise in analyzing therapeutic progress and emotions (Norman et al., 2020; Tanana et al., 2021), and even predicting outcomes (Van Le et al., 2018). These advances highlight the potential of NLP-based tools, including generative AI models like ChatGPT, to enhance therapeutic processes by supporting clinicians in tasks such as analyzing patient narratives and providing structured feedback to improve case conceptualization. Large Language Models (LLMs) like ChatGPT represent a significant evolution in NLP, with their ability to generate contextually appropriate and coherent responses (Bommasani et al., 2021; Brown et al., 2014). Unlike traditional NLP tools, LLMs rely on extensive training datasets and transformer architectures to produce human-like reasoning and adapt to a variety of tasks. In clinical contexts, LLMs hold promise for aiding in complex analyses, such as interpreting patient statements or aligning them with theoretical frameworks (McCoy et al., 2024). Their ability to provide real-time feedback on clinical constructs, such as therapeutic narratives or case formulations, further enhances their relevance in psychotherapy. However, their performance is not without limitations. Despite their versatility, LLMs often rely on pattern recognition rather than true comprehension, which can lead to errors in nuanced scenarios (Jones & Bergen, 2023).

ChatGPT, as a prominent example of an LLM developed by OpenAI in 2015, generates responses that closely resemble human language in naturalness and coherence (Khurana et al., 2022; Liu, 2024). While its ability to simulate human dialogue is notable, its application in clinical decision-making or theoretical analysis remains largely unexplored. Concerns persist about its capacity to process abstract psychological concepts and manage complex emotional constructs effectively (Lee et al., 2023). Understanding these capabilities and limitations is essential to evaluate how tools like ChatGPT can complement clinical practice without undermining the clinician-patient relationship (Wong, 2024).

Case conceptualization, which serves as a bridge between a client's concerns and therapeutic strategies, is a critical aspect of psychotherapy (John & Segal, 2015). Among existing models, LIBET (*Life themes and plans Implicated in Biases: Elicitation and Treatment*; Sassaroli et al., 2022, 2023) is a case formulation technique grounded in the diathesis-stress model and rooted in the cognitive-behavioral tradition (Beck, 2011). It is organized along two dimensions: 'life themes' and 'semi-adaptive plans' (Sassaroli et al., 2022). Life themes are emotional sensitivities and core beliefs shaped by early experiences and relational environments, each grounded in a distinct emotional fundation. They include *Unloved* (associated with sadness and emotional detachment, often stemming from perceived abandonment), *Threatened* (rooted in fear and a lack of safety, typically arising in unstable or unpredictable environments), and *Unworthy* (characterized by shame, inferiority, and self-disgust, often linked to critical or rejecting relational dynamics). Semi-adaptive plans are rigid strategies that patients develop to identify and prevent the risk of emotional pain or adverse situations linked to the life-theme. These include the *Prudential* plan (avoidance of threats through passive behaviors or mental avoidance, linked to anxiety and shame), the *Prescriptive* plan (control through compulsive worrying and over-controlling behaviors, tied to anxiety and guilt), and the *Immunizing* plan (cognitive-affective manipulation or self-reward, linked to anger and desire). These plans are regulated by a higher-order mechanism known as metacontrol, which governs how flexibly individuals manage conflicting internal needs and maintain coherence in their responses. From a psychopathological standpoint, the LIBET model suggests that suffering arises not from the mere presence of emotional sensitivities and coping strategies—such as life themes and semi-adaptive plans—but from the rigidity of their use and the inflexibility of metacognitive control mechanisms. Importantly, these constructs are considered inherent across individuals, making LIBET-based interpretations applicable even in non-clinical narratives. See Table 1 for further detailed definitions. These constructs are thought to offer a clear framework for assessing the depth and accuracy of ChatGPT's clinical reasoning, due to its structured approach.

Existing research highlights the link between successful therapeutic outcomes and effective case conceptualization, yet the potential of AI-generated feedback in this context remains underexplored (John & Segal, 2015). Clinical feedback, traditionally delivered either through clinical supervision or via systematic client-informed mechanisms, enhances therapeutic alliance, improves clinical competence, and contributes to symptom reduction by providing structured, real-time insights during therapy (Bambling et al., 2006; de Jong et al., 2021, 2025; Weck et al., 2016). These findings lay the groundwork for investigating how AI-generated feedback could similarly assist clinicians in refining case conceptualization and therapeutic decision-making. Moreover, ethical and practical concerns about the integration of AI in psychotherapy—such as the risk of over-reliance on AI and its ability to navigate sensitive emotional content—demand further investigation (Pham et al., 2022; Wong, 2024).

This exploratory study aims to qualitatively assess ChatGPT-4's ability to apply the LIBET case formulation model (Sassaroli et al., 2023) by analyzing its feedback on anonymized interview transcripts. Specifically, the study investigates whether the chatbot can accurately identify and interpret core psychological constructs—life themes and semi-adaptive plans—while adhering to theoretical principles. At the same time, the study examines the nature of recurring errors and reasoning patterns in the chatbot's output in order to evaluate its clinical applicability and areas in need of refinement. The study is hence guided by two central research questions:

(1) Does ChatGPT-4 generate feedback that reflects a coherent use of the LIBET model's constructs, based on the content of participant narratives?

(2) What types of interpretative errors or reasoning patterns emerge in its output, and how might these inform its potential clinical use?

The findings are intended to provide insights into this model's capabilities and limitations, and to offer practical recommendations for its future refinement and integration into clinical support systems and reasoning processes.

Table 1. *Definitions of Life Themes and Semi-Adaptive Plans in the LIBET Model.*

Construct	Categories	Description
Life Themes		Core self-beliefs shaped by early experiences and relational environments. They reflect fundamental perceptions of oneself in relation to safety, value, emotional connection, and worth, influencing emotional states and patterns of behavior.
	Unloved	Feeling of loss, emotional detachment or abandonment, and futility, linked to an emotional state of depressive sadness.
	Threatened and Inadequate	Feeling of danger and lack of personal safety and protection guaranteed materially and affectively by significant and reliable figures, linked to an emotional state of fear or panic.
	Unworthy and Inadequate	Feeling of exclusion, inferiority, and contempt towards oneself, linked to an emotional state of shame and guilt.
Semi-Adaptive Plans		Coping strategies developed to manage emotional pain or adverse situations. While they provide temporary relief, these patterns often hinder growth and reinforce maladaptive behaviors, linking emotions like anxiety, guilt, anger, or desire to specific responses.
	Prudential	Avoidance of aversive and threatening stimulations, both behaviorally and mentally, leading to failure in developing exploratory and constructive aspects of existence, linked to anxiety and/or shame.
	Prescriptive	Attempts to control, prevent, or resolve adverse stimuli through compulsive worrying and controlling behaviors, linked to anxiety and/or guilt.
	Immunizing	Exclusion of painful states through emotional manipulation, including intense interpersonal anger or self-rewarding behaviors (e.g., substance use), linked to anger or desire.

Note. Construct: LIBET main constructs; Components: sub-dictions of main constructs; Description: a brief explanation of each component.

Methods

This study received ethical approval from the Ethics Committee of Sigmund Freud University (Reference: PD48HK50C37GJ490824).

Participants

A purposive sample of ten adult participants (5 female, 5 male; age range: 22–41 years) was selected for this study. Inclusion criteria required participants to be over 18 years of age, native Italian speakers, and to have no self-reported history of psychological disorders or formal training in psychotherapy or clinical psychology. These criteria were designed to minimize potential biases related to prior theoretical knowledge or lived clinical experiences that could influence narrative content. The use of a non-clinical sample also reflected ethical constraints associated with inputting clinical data into commercial AI tools such as ChatGPT. This exploratory design allowed for a preliminary evaluation of the model's interpretive processes in a controlled, low-risk context. Participants were recruited in January 2024 through word-of-mouth and social media, after receiving a detailed explanation of the study's objectives and procedures. All participants provided informed consent in accordance with ethical guidelines.

Data Collection

Semi-structured interviews, grounded in the LIBET model and aligned with its manual guidelines (Sassaroli et al., 2023), were conducted between January and February 2024. All participants provided informed consent prior to participation. Interviews were conducted by two trained researchers (M.B. & L.P.) who had received standardized instruction on the interview protocol, including the duration, phrasing, and sequence of questions, to ensure procedural consistency. Demographic information was collected at the beginning of each session.

The interviews were conducted via Microsoft Teams, audio-recorded, and subsequently transcribed verbatim. Transcripts were manually reviewed by three researchers (M.B., D.B. & L.P.) to correct transcription errors (e.g., word omissions, misspellings) and ensure fidelity to the original content. Each participant contributed a single semi-structured interview, resulting in a total of ten transcripts and a total of ten feedbacks (five for each analysis).

To protect participant confidentiality, all personally identifiable information was removed during transcription. Sensitive linguistic content—such as names, locations, dates, sexual orientation and specific age references—was replaced with neutral placeholders (e.g., [name], [location], [date], [age]). Dialogue exchanges were consistently coded using the labels “I” (Interviewer) and “P” (Participant) to support traceability during thematic analysis.

Design of the ChatGPT Interaction

The GPT-4 model (via the ChatGPT Premium subscription) was deployed via the MyGPT interface to support the analysis of interview transcripts. The chatbot, titled “LIBET-Chat Interviste”, was configured in December 2023 and subsequently used between January and February 2024 (OpenAI, 2024). Its objective was to analyze Italian-language transcripts of semi-structured interviews with non-clinical participants and generate interpretative feedback based on the LIBET model (Sassaroli et al., 2023), specifically focusing on the identification of life themes and semi-adaptive plans.

In the Custom GPT configuration phase, a personalized MyGPT chatbot was created using OpenAI’s GPT Builder. It was configured through tailored background instructions written in Italian, including a thorough description of the LIBET model and an uploaded PDF of the official manual to ensure theoretical accuracy. Although no model weights were altered, the chatbot was personalized through carefully designed prompt instructions and domain-specific materials. The instructions explicitly outlined the chatbot’s intended function to offer structured feedback on interview content by applying the LIBET framework while maintaining formal psychological terminology and avoiding irrelevant or speculative content (see Appendix A for the configuration exchange).

Moreover, the interaction protocol required the chatbot to adopt a formal tone and communicate in Italian, addressing the psychologist-interviewer (“I”) while referring to the participant (“P”) in the third person. Emphasis was placed on the logical reasoning behind each interpretation, and the use of bullet points or lists was explicitly discouraged in favor of paragraph-based responses. These design elements aimed to elicit feedback with a consistent structure and linguistic form across different transcripts, allowing for a more focused evaluation of content accuracy regarding the LIBET constructs, rather than prompting diagnostic reasoning.

Prompt Engineering

The prompt design followed the guidelines available on the official OpenAI documentation (Lee et al., 2023; OpenAI Platform, n.d.) and was developed using prompt engineering techniques tailored to ensure both theoretical accuracy and linguistic clarity. The aim was to instruct the chatbot to generate consistent, theory-based outputs aligned with the LIBET framework (Sassaroli et al., 2023). Prompt construction was guided by key principles including explicit task definition, background context, role assignment, and output constraints (e.g., word count, tone, language). Specifically, each prompt included: (i) a cleaned and anonymized transcript of an interview with a non-clinical participant, (ii) a detailed description of the theoretical model (LIBET), including all three possible life themes, and (iii) a direct instruction to hypothesize the participant’s life theme (or semi-adaptive plans) based on the interview content.

The chatbot was asked to articulate its reasoning process prior to providing a conclusion, emphasizing the analytical path over the final judgment. The language of interaction was set to Italian, with explicit instructions to maintain a formal tone, avoid bullet points, and adopt psychologically accurate terminology. A full example of the original prompt (both in Italian and translated in English) is available in the Appendix B.

This procedure was applied consistently across all interviews and across both constructs investigated. Prompt design played a crucial role in optimizing output quality and interpretative consistency, in line with recent recommendations for responsible and transparent AI use in mental health research (Beg, 2025).

Thematic Analysis

Two rounds of reflexive thematic analysis were conducted to identify patterns in the data, following Braun and Clarke's (2006, 2021b) six-phase framework. The approach was primarily inductive, but was theoretically informed by the LIBET model (Sassaroli et al., 2023), which provided a conceptual lens for interpreting and organizing the themes. In line with the assumptions of reflexive thematic analysis and the theoretical insight focus, data saturation was not considered a methodological goal (Braun & Clarke, 2021a). We acknowledge that the researchers' backgrounds may have influenced the interpretation process. The combination of clinical and academic perspectives—ranging from strong familiarity with the LIBET model to more exploratory engagement—was intended to support both theoretical alignment and critical distance. The dual role of interviewer-analyst was discussed throughout the analytic process, and potential bias was mitigated by the fact that one of the two primary coders had not conducted the interviews, allowing for partial separation between data collection and analysis.

Once ChatGPT's feedback was collected for all transcripts, two researchers independently reviewed each response in full. During this first phase, they read the texts multiple times, paying attention to recurring patterns in how the chatbot processed and expressed LIBET-related constructs. Rather than applying a fixed coding frame, they allowed categories to emerge inductively, while using the LIBET manual as a conceptual guide to interpret and label the themes.

In a second round of independent reading, each researcher annotated the texts with preliminary observations and theoretical reflections. These annotations were then re-examined individually in a subsequent phase, during which each researcher compared their own notes and began identifying recurring features. Based on this process, each researcher independently proposed provisional thematic categories.

To enhance reliability and reduce subjective bias, a third researcher [blinded for review] independently reviewed the emerging themes and subthemes, qualitatively assessing the conceptual overlap between the two initial coders as a check to ensure thematic coherence and agreement. To assess inter-rater agreement, Cohen's Kappa (κ) was also calculated for each thematic category. For some categories, κ values were close to 0 despite high observed agreement (e.g., 90%), due to strongly unbalanced coding distributions. This reflects a known limitation of κ in small and skewed samples. Therefore, raw agreement percentages were also reported to provide a more accurate representation of coder consistency (see Table 2).

Table 2. *Inter-Rater Agreement for Thematic Categories Identified in ChatGPT Responses.*

Thematic Category	1	2	3	4
κ Cohen	0.375	0	0	0.231
% Agreement	80%	90%	100%	80%

Note. κ Cohen = Cohen's Kappa coefficient, measuring inter-rater reliability adjusted for chance agreement. Thematic Category 1 = Limitations in Abstraction and Interpretative Barriers; 2 = A Consistent Structure and Content Organization; 3 = Hypothesis-Driven Reasoning: From Hypothetical Language to Practical Examples; 4 = Expressed Theory and Specific Language Adhering to LIBET; Components: key elements observed within each theme; Description: a brief explanation of each component's relevance or implications.

Following this, all three researchers collaboratively refined the thematic structure through joint discussion, aiming to establish convergence and resolve discrepancies. They reviewed and named the themes and subthemes to ensure conceptual clarity, theoretical alignment (including consultation of relevant literature), and representativeness of the data. The final thematic map was iteratively reviewed until full consensus was reached, and illustrative quotes from ChatGPT's responses were selected to support each theme.

After the completion of both analyses (focused respectively on life themes and semi-adaptive plans), the three researchers jointly compared the resulting thematic structures to assess their consistency and potential integration. This final step supported the identification of overarching categories and ensured a coherent thematic framework across both constructs. A detailed description of the thematic analysis process, along with supplementary materials, is available at the following OSF project page: https://osf.io/bqt6w.

Results

Following the completion of the two separate thematic analyses—one focused on life themes, the other on semi-adaptive plans—the researchers conducted a joint comparison to identify overarching patterns across both sets of data. This process revealed a significant convergence in interpretative features, allowing the results to be synthesized into four main themes.

The first theme, Limitations in Abstraction, underscores ChatGPT's interpretative challenges and barriers, leading to imprecise content and misaligned feedback on both the clinical constructs. It emerged, for example, when ChatGPT inferred a life theme of Unworthiness based on behavioral patterns like *asking a sibling for help with a bill,* yet referenced only emotions such as *frustration* and *nervousness,* without linking them to deeper affective states like shame or inadequacy. Similarly, in analyzing semi-adaptive plans, the model described avoidance and control behaviors—such as *sacrificing personal needs* or *rationalizing the partner's emotional reactions*—but often failed to explicitly connect these to underlying emotions like anxiety or guilt.

The second theme, Structure and Form, highlights systematic patterns in ChatGPT's responses, characterized by a consistent organization of the content and subtle variations in structural choices. The third theme, Hypothesis-Driven Reasoning, reflects ChatGPT's research-oriented approach, where it used probabilistic reasoning to test hypotheses, referencing the LIBET model and illustrating constructs through practical examples grounded in real-life scenarios. Finally, the theme Expressed Theory and Individual Plans reflects ChatGPT's application of theoretical models. It adhered closely to LIBET's principles and incorporated individual control dynamics, metacontrol, and the crystallization of psychopathological frameworks. See Table 3 for a detailed summary and brief descriptions of the components within each thematic category.

Limitations in Abstraction and Interpretative Barriers

The analysis revealed limitations in the model's ability to map theoretical constructs accurately, leading to vague or overlapping interpretations, especially evident in its handling of life themes. A recurring issue was the AI's difficulty in distinguishing between life themes and semi-adaptive plans, leading to frequent misinterpretations.

In its analysis of life themes, ChatGPT's output tended to prioritize current coping behaviors over the core emotional vulnerabilities that define the construct. For instance, it emphasized behaviors such as *concerns about self-worth* rather than exploring the emotional history underpinning the themes, resulting in incomplete interpretations. Moreover, the system exhibited limitations in processing emotional sensitivity inherent to life themes, often focusing on surface-level behaviors rather than deeper emotional dynamics: "Self-criticism for forgetfulness and distraction, along with a tendency to diminish one's ability to independently manage challenges, are typical characteristics of this theme, which includes feelings of inferiority and inadequacy, as well as a struggle with self-worth and competence."

Regarding semi-adaptive plans, ChatGPT often demonstrates reduced accuracy in its analysis, with misinterpretations of the true function of certain behaviors within the broader psychological framework. For example, it inaccurately linked aspects of the prescriptive and prudential plans, such as when it suggested: "The prescriptive plan is evident in how they manage their relationship with their partner, trying to control and prevent conflicts, as seen in their attempts to rationalize and normalize their partner's emotional reactions."

Additionally, ChatGPT's analyses often lacked depth, particularly in explaining how control strategies interacted with deeper psychological vulnerabilities. This limitation was compounded by its tendency to overlook participants' emotional history and rely on repetitive reasoning, which reduced the clinical relevance of its feedback. For example: "The desire for independence and active management of one's career, combined with feelings of loneliness, worthlessness, and the unfulfilled search for meaningful relationships, suggest that the 'Dislove and Inadequacy' life theme better represents P's current emotional and cognitive experience."

Table 3. *Key Observations and Brief Descriptions of Thematic Components in ChatGPT's Feedbacks.*

Theme	Components	Description
1	Limited Emotional Depth	Mentions few emotions, primarily tied to the present. It also fails to bridge deeper emotional dynamics to the self-cognitions that define life themes (e.g., *emptiness and sadness* tied to *I am alone*)
	Misinterpreted Behaviors	Coping behaviors were often mistaken for emotional vulnerabilities, reducing interpretative accuracy (e.g., *I struggle to be enough, hence I fear not to be enough*)
	Lack of Depth	Limited ability to capture the emotional sensitivity and the role of emotional history inherent in life themes
	Distinguishing Constructs	Difficulty in differentiating some constructs (e.g., *Prudential control* and *Active prescriptive control*), leading to conceptual overlaps
	Repetitive Reasoning	Overlooked some in-depth information, often leading to repetitive and surface-level interpretations
2	Summary Conclusions	Consistently provides summary conclusions that reiterate key concepts, occasionally noting missing elements for hypothesis development or suggesting potential interventions
	Formatting Clarity	Clear formatting (e.g., capitalization of terms) to distinguish constructs like plans and themes
	Progressive Structure	Begins with a general overview before delving into specific content, maintaining logical consistency across narrative elements
	Structured Feedback	Consistent use of sections (introduction, reasoning, summary) to organize responses
3	Probabilistic Reasoning	Used cautious, hypothetical language (e.g., *This may suggest*) to propose interpretations
	Theoretical Alignment	Aligned hypotheses with theoretical constructs to maintain coherence with the LIBET framework
	Cautious Proposals	Avoided definitive conclusions, favoring a careful, step-by-step reasoning process
	Grounded Examples	Transitioned from abstract hypotheses to real-life (and transcript-derived) examples, enhancing practical relevance
4	Avoidance Across Plans	Recognizes that all semi-adaptive plans function as forms of avoidance of life themes, avoiding the misinterpretation of *avoidance* as exclusive to the prudential plan
	Linking Narratives to Theory	Demonstrates coherence in connecting participants' descriptions of experiences and emotions to LIBET constructs
	Dominant Plans	Successfully identified key semi-adaptive plans within participant narratives
	Control Across Domains	Extends the concept of control to various contexts, including relational dynamics and performance, demonstrating flexibility in applying the construct
	Specialized Terminology	Effective use of clinical and LIBET-specific terms such as *inadequate*, *metacontrol* and *conflicting needs*
	Hypercontrol and Rigidity	Recognized hypercontrol strategies and their rigidity, including their potential for emotional breakdowns

Note. Theme: the overarching thematic category identified in the analysis; 1 = Limitations in Abstraction and Interpretative Barriers; 2 = A Consistent Structure and Content Organization; 3 = Hypothesis-Driven Reasoning: From Hypothetical Language to Practical Examples; 4 = Expressed Theory and Specific Language Adhering to LIBET; Components: key elements observed within each theme; Description: a brief explanation of each component's relevance or implications.

A Consistent Structure and Content Organization

ChatGPT's output showed a consistent and methodical approach in organizing its feedback, applying a clear structure to its analyses. It reliably divided feedback into clear sections, beginning with an introduction, followed by segmented reasoning, and concluding with a summary to reinforce key hypotheses and main findings, such as: "In conclusion, […] the 'Dislove and Inadequacy' life theme is more representative of P's current emotional and cognitive experience."

It also employed formatting strategies, such as capitalizing key terms, to clarify shifts in the discussion: "First, we can hypothesize the adoption of the Prudential Plan […] Second, the Prescriptive Plan […]"

The AI's texts maintained a structured format throughout its analyses, often starting with a brief hypothesis before organizing its observations. In some cases, it first provided an analysis of the content before introducing specific terms, while in others, it moved directly to key identifications: “In the interview transcript […] it is possible to hypothesize that the participant (P) exhibits the prescriptive semi-adaptive plan.”

Its language was formal and professional throughout. Occasionally, therapeutic interventions were suggested or the absence of certain constructs was noted, as in the following example: “There seems to be less evidence of the immunizing plan […] however, this aspect may require deeper analysis.”

Hypothesis-Driven Reasoning: From Hypothetical Language to Practical Examples

ChatGPT demonstrated a hypothesis-driven approach to analyze the main constructs.

The AI system consistently employed probabilistic language, producing plausible hypotheses grounded in emotional, psychological, and behavioral data from the transcripts. Rather than generating definitive conclusions, theoretical concepts were carefully aligned with specific aspects of the interview content: “This internal contradiction seems to be rooted in […] The deep sense of insecurity and fear of inadequacy […] seem to reflect a life theme of […].”

Similarly, interpretations were framed cautiously, as shown in another example: “This may suggest that their life theme includes a strong need for control and competence.”

Observable data and the transcripts’ content were often referenced to support hypotheses, as in this example: ”During the interview, P expresses dissatisfaction and a tendency to frequently change work and residential environments, highlighting a continuous search for novelty and stimulation.”

These observations were coherently aligned with theoretical constructs, ensuring a grounded and systematic approach: “This internal contradiction seems to be rooted in […] fear of inadequacy.”

However, when narratives were less detailed or ambiguous, ChatGPT’s reasoning became less specific, often resulting in generalized conclusions: “This is evident from their account of how they handle complex work situations or potentially problematic social interactions, where an avoidance attitude prevails.”

Expressed Theory and Specific Language Adhering to LIBET

ChatGPT’s output reflected a strong ability to apply LIBET theory, effectively using precise psychological terminology and connecting theoretical constructs to participants’ narratives. It frequently employed terms such as *inadequacy* and *expectations,* while referencing cognitive mechanisms like metacontrol, attention, and conflicting needs, thought as the internal psychological tension that arises when an individual experiences competing desires or goals that are in conflict with each other: “P also exhibits resistance to engaging in tasks they perceive as difficult, preferring to avoid effort rather than risk failure.”

Subjects’ narratives were adeptly linked to emotional struggles, particularly issues of self-perception: “This internal contradiction seems to be rooted in a distorted self-perception, where P sees themselves as inadequate despite external evidence to the contrary.”

In analyzing semi-adaptive plans, the model’s responses captured elements of control dynamics and pursued its hypotheses based on observed behaviors: “The prescriptive plan is evident in how the participant manages the relationship with their partner, trying to control and prevent conflicts.”

The identification of dominant plans within narratives was also effective. Clinically speaking, individuals may exhibit the use of multiple strategies related to different semi-adaptive plans. They can indeed develop one or two plans that they use predominantly, meaning that these plans are more pervasive in their lives. The concept of dominant plan here refers to the semi-adaptive strategy a person primarily uses to cope with life challenges: “[…] P has predominantly activated the prudential and prescriptive plans.”

A recurring sub-content was Metacontrol and Rigidity. Metacontrol is a cognitive regulation process that involves higher-level strategies for regulating thoughts and behaviors (Sassaroli et al., 2023). Rigidity in metacontrol emerges when these strategies become inflexible, reducing adaptability. This often leads to overly adhere to fixed rules or cognitive scripts, even in contexts where flexibility would be beneficial. The model’s output included references to hypercontrol strategies as central to participants’ coping mechanisms: “In the Prescriptive Plan, the participant clearly demonstrates the use of hypercontrol strategies to avoid future adverse situations…”

Further mentions of plan rigidity appeared in ChatGPT's responses, noting their potential for breakdown and the emotional frustration when they fail. This observation refers to the concept of plan invalidation, that occurs when an individual's semi-adaptive plan fails to achieve its intended goals, and hence to be adaptive. For example, the invalidation of the prescriptive plan occurs when an individual rigidly follows rules or norms to gain approval through external or internal validation but fails to achieve the desired result. This causes emotional frustration and reinforces a rigid cognitive regulation. This sometimes leads to symptomatic outcomes. "The interviewee describes a constant and disproportionate commitment to controlling their performance and maintaining relationships, even at the expense of their own happiness and well-being."

Discussion

This study explored the applicability of ChatGPT within a psychological case formulation model—LIBET (Sassaroli et al., 2022, 2023)—by analyzing its ability to identify life themes and semi-adaptive plans from anonymized interviews. The results indicate that ChatGPT demonstrated a consistent capacity to apply key LIBET constructs in its feedback, using appropriate psychological terminology such as metacontrol, and conflicting needs. Its responses followed a structured, hypothesis-driven approach, often using cautious, probabilistic language (e.g., *this may suggest*), which helped avoid overinterpretation or premature conclusions (Lee et al., 2023).

Despite this promising performance, ChatGPT showed limitations in abstraction and interpretative depth—particularly when reasoning about emotionally rooted constructs like life themes. These difficulties are consistent with prior research suggesting that LLMs, while capable of generating coherent and contextually appropriate text, often lack the nuanced comprehension needed for deeper emotional or clinical interpretation (Beg et al., 2024; Pham et al., 2022).

The distinction between surface-level behaviors and the underlying emotional vulnerabilities—central to the LIBET model—was particularly challenging for ChatGPT. As highlighted in the analyses, the model occasionally conflated semi-adaptive plans with life themes, demonstrating a superficial integration of emotional dynamics. These limitations underscore the need for ongoing refinement and careful consideration when using LLMs for clinical applications (Beg, 2025; Stade et al., 2024).

Nevertheless, the model's structured reasoning and adherence to theoretical language offer a foundation for future implementation in clinical reasoning support, particularly in tasks involving pattern detection, hypothesis generation, and supervision support (Kjell et al., 2024; Lee et al., 2023; Stade et al., 2024).

Error Analysis and AI Reasoning

While this study did not quantify the frequency or typology of ChatGPT's errors, several recurring patterns emerged from the thematic analyses. The most prominent issue concerned the conflation of surface-level coping strategies with underlying emotional vulnerabilities. For example, behaviors associated with semi-adaptive plans were often interpreted as indicative of a life theme, resulting in blurred distinctions between constructs. This interpretative overlap reflects a form of conceptual misattribution and may limit the clinical validity of AI-generated feedback.

Another type of error observed involved emotional misalignment: the model occasionally identified emotions such as sadness, fear, or insecurity without adequately linking them to the broader psychological context or developmental history that defines life themes in the LIBET model (Sassaroli et al., 2022). This reflects a limited capacity to contextualize emotional content within broader psychological frameworks, consistent with emerging discussions on the challenges of applying AI tools to emotionally nuanced material and clinical decision-making (Beg et al., 2024; Pham et al., 2022), difficulties that may be further exacerbated when the target constructs are poorly operationalized or show conceptual overlap.

Despite these challenges, ChatGPT consistently employed a cautious and hypothesis-driven style of reasoning. It relied on probabilistic language—such as *this may suggest* or *it is possible that*—to articulate interpretations. While such phrasing might reduce clarity in certain clinical contexts, it can also function as a safeguard against overconfidence and premature conclusions, especially in early-stage reasoning. This linguistic strategy aligns with ethical recommendations for mitigating risks of overinterpretation in AI-generated feedback (Beg, 2025; Lee et al., 2023), and may support reflective clinical thinking where nuance and uncertainty are inherent. However, some outputs included instances of hallucinations—confident but incorrect assertions unsupported by the transcript data or LIBET theory. These errors, while not systematically tracked in the present study, pose a risk in therapeutic settings where interpretative accuracy is critical. As such, future studies should consider including a quantitative classification of error types (e.g., conceptual misinterpretation, emotional misattribution, unsupported claims), alongside linguistic analyses of AI outputs to better assess reliability and clinical applicability.

Theoretical and Clinical Implications

The findings of this study suggest that while ChatGPT can apply psychological theory in a structured and coherent manner, it struggles with interpreting emotionally embedded and abstract constructs such as life themes. In the LIBET model, life themes reflect deeply rooted emotional sensitivities shaped by early relational experiences (Sassaroli et al., 2022). These themes are not easily accessible through explicit verbalization and often manifest through implicit emotional cues.

From a broader theoretical perspective, these themes might also be influenced by embodied processes, where early sensorimotor interactions with caregivers—such as touch and proximity-seeking—play a foundational role in shaping attachment schemas and emotional patterns (Beckes et al., 2015). Similarly, early environmental interactions integrate physical and emotional cues, as shown in cognitive and linguistic development (Wellsby & Pexman, 2014), which may underpin the abstract and relational nature of life themes, and influence abstract cognitive representations, including those related to self-worth, safety, and connection. This embodied foundation may help explain why life themes—being also anchored in non-verbal, experiential memory—pose greater interpretative challenges for LLMs like ChatGPT, which rely heavily on linguistic patterns and explicit content. In contrast, semi-adaptive plans, which are more action-oriented and often described in concrete behavioral terms, appear more accessible to the model.

Furthermore, the chatbot's responses tended to conflate behavioral coping strategies with emotional core beliefs, which illustrates a superficial grasp of the dynamic relationships within the LIBET framework. This pattern suggests that, while the model can produce coherent outputs through linguistic alignment with theoretical terminology, it may do so without fully integrating the relational and emotional underpinnings of the constructs involved. While ChatGPT's responses were structured and aligned with theoretical terminology, they may ultimately reflect advanced statistical associations rather than a conceptual representation of psychological constructs.

These limitations notwithstanding, the model's use of structured reasoning and probabilistic language still offers potential clinical utility. It could be leveraged as a support tool in psychotherapy training or supervision contexts, where it may help practitioners reflect on patterns and hypotheses within client narratives without replacing clinical judgment (Stade et al., 2024).

Ethical and Practical Considerations

The integration of generative AI into mental health research holds both promise and risk. On one hand, tools like ChatGPT offer scalable support for analyzing psychological material, assisting clinicians in reflective processes and improving access to structured feedback. On the other hand, recent scholarship has highlighted critical concerns regarding accuracy, bias, and the erosion of human judgment when AI is deployed without sufficient oversight (Beg, 2025).

A central ethical issue is the phenomenon of "AI hallucinations," where language models generate content that is syntactically coherent yet factually incorrect or unsupported—an issue observed in this study as well (Lee et al., 2023). As Beg (2025) notes, such inaccuracies can undermine the reliability of AI-supported research, especially in high-stakes fields such as healthcare. Additionally, the "black box" nature of AI models obscures their internal reasoning processes, limiting transparency and accountability (Jones & Bergen, 2023).

Over-reliance on AI also raises concerns about the dilution of critical thinking and intellectual responsibility. As Beg (2025) emphasizes, the use of AI should not replace human reasoning, but rather augment it, maintaining the central role of human expertise in academic and clinical decision-making. In psychological research, this is particularly crucial: interpreting emotionally nuanced or abstract constructs—such as those explored in the LIBET model—requires contextual sensitivity that AI alone cannot provide.

Moreover, data privacy must be carefully protected. Although this study worked with anonymized, non-clinical material, future implementations involving sensitive clinical data will require rigorous safeguards. According to Beg (2025), ethical frameworks must ensure transparent disclosure of AI use, responsible data handling, and adherence to international guidelines (e.g., COPE, CONSORT-AI, ICMR).

Ultimately, responsible integration of AI in mental health research must be grounded in transparency, methodological rigor, and human oversight. Researchers should clearly report the scope and limits of AI involvement, validate outputs through independent verification, and remain accountable for the quality and integrity of their work (Beg, 2025; Lee et al., 2023).

Limitations

Despite its promising findings, this study presents several limitations. The use of a small, non-clinical sample restricts generalizability, particularly regarding clinical applicability. This choice was guided both by ethical constraints—given the risks of using clinical data with commercial AI models—and by the need to test AI performance in a controlled setting.

However, the LIBET model posits that life themes and semi-adaptive plans are core psychological structures, present across both clinical and non-clinical individuals (Sassaroli et al., 2022). What distinguishes clinical suffering is the rigidity of metacognitive regulation (i.e., the use of my plan is a need vs is an option), not the mere presence of these constructs. Recent findings by Offredi et al. (2025) confirm this distinction: while both groups endorsed similar themes and plans, only the clinical group showed significantly higher dysfunction in metacognitive control. These insights support the theoretical rationale for using non-clinical data in this preliminary phase, while underscoring the importance of future studies involving clinical samples to evaluate AI's interpretative accuracy under more complex conditions. Moreover, ChatGPT-4's outputs were based solely on the theoretical material provided through custom instructions and not on any prior embedded knowledge of the LIBET model, as its training data cutoff was September 2023. This further supports the rationale for using a controlled, non-clinical dataset in this preliminary phase, while highlighting the need for more advanced customization methods and external validation in future studies.

Second, the study relied exclusively on Italian-language transcripts, which may have affected ChatGPT's performance (Wendler et al., 2024). Although the model demonstrated adequate comprehension and terminology use, LLMs are typically optimized for English, and subtle nuances of meaning may be lost in translation or less accessible in multilingual contexts.

Some of the research team conducted both the interviews and the thematic analysis. Although independent rounds of coding and review by an external coder were implemented to reduce bias, the dual role may have influenced interpretation. Additionally, while raw agreement between coders was high, Cohen's κ values were low—reflecting known limitations of this metric in small and unbalanced datasets. This reduces the interpretability of inter-rater agreement and highlights the need for future studies to involve larger, independent coding teams and to consider complementary reliability measures.

Finally, although various interpretative limitations—such as hallucinations and conceptual inaccuracies—were identified and illustrated (see Results), no formal linguistic or quantitative error analysis was conducted. This choice aligns with the principles of reflexive thematic analysis (Braun & Clarke, 2006, 2021b), which discourage frequency-based coding in meaning-oriented research. However, the absence of structured quantification limits the ability to assess the prevalence and potential impact of such phenomena. Future studies should include systematic analyses to evaluate the reliability and clinical relevance of AI-generated interpretations.

Future Directions and Recommendations

Building on the current findings, future research should further investigate the use of AI tools as clinical decision-support systems rather than substitutes for therapeutic reasoning. In this study, ChatGPT-4 demonstrated a capacity to provide structured conceptual interpretations grounded in a psychological model, suggesting its potential utility as a reflective aid for clinicians during case formulation or supervision, particularly in structured theoretical approaches like LIBET. As emphasized by Beg and Verma (2024), successful implementation of AI-assisted interventions depends not only on technological performance but also on the quality of the therapeutic alliance they help foster. Improving user engagement and the digital therapeutic alliance is therefore critical for ensuring the effectiveness of AI-supported interventions, as consistently highlighted in recent reviews (Beg & Verma, 2024). While much of the literature focuses on digital therapeutic alliance with clients, future research should also explore the concept of a digital Supervisory Working Alliance (Enlow et al., 2019), that is, how AI-supported feedback might enhance a clinician's reflective process, promote theoretical fidelity, and facilitate supervision, especially in early-career therapists. Building on this perspective, AI systems must indeed be designed not only for accuracy but also for empathic alignment, ensuring that the interaction remains supportive, non-judgmental, and therapeutically relevant. This calls for user-centered design strategies that simulate empathic responses while maintaining psychological coherence.

To build on this relational foundation, future research should also prioritize the refinement of prompt engineering and potentially structured tuning approaches (OpenAI Platform, n.d.). As shown in this study, ChatGPT-4 was able to provide theoretically informed responses when supplied with structured background instructions and clearly formulated tasks. However, the observed limitations suggest that more precise and layered prompting techniques could enhance the model's clinical reasoning. Given the model's sensitivity to linguistic structure and instruction hierarchy, tailored prompting could help reduce conceptual misattributions and increase the specificity of outputs. A further step would assess whether AI systems can detect not only the presence of semi-adaptive plans but also their degree of rigidity, a central metacognitive feature in the LIBET framework. This metacognitive feature—core to the model—could offer a more nuanced understanding of dysfunctional patterns and enhance the clinical applicability of automated feedback. This approach aligns with the broader goal of designing AI-assisted tools that support—not replace—clinical judgment.

In practical terms, future refinements could also involve testing shorter and more targeted excerpts of clinical transcripts to isolate specific interpretative challenges, or comparing outputs across different LLMs, including reasoning-optimized models. Additionally, the development of a training set composed of annotated transcripts and expert-coded LIBET constructs could support the development of structured fine-tuning pipelines for AI outputs. To move in this direction, it is essential to assess the capabilities and limits of each specific model before deploying it in assistive or collaborative roles (Stade et al., 2024).

Future research might also adopt alternative qualitative strategies—such as comparative content analysis or typological saturation approaches—to investigate whether AI-generated feedback converges across larger datasets or varied contexts. While methodologically distinct from reflexive thematic analysis (Braun & Clarke, 2021a, 2021b), they could provide useful insight into the consistency, variability, and generalizability of interpretative patterns in LLMs, particularly in relation to their initial coding behavior and saturation thresholds (De Paoli & Mathis, 2025).

Moreover, Beg et al. (2024) underscore the need for ethical safeguards in the deployment of AI mental health tools, including data protection, informed consent, and bias mitigation. These considerations are particularly important when extending such tools to clinical populations. Although this study used a non-clinical sample for ethical reasons, the approach could be adapted for clinical contexts—especially in supervision or team-based case discussions—where structured feedback generated through validated theoretical frameworks could enhance insight and intersubjective consistency. To responsibly integrate such tools into psychotherapeutic practice, it is essential to ensure transparent reporting of AI use, critical evaluation of its outputs, and continuous human oversight. This also entails safeguarding informed consent, protecting data privacy, and mitigating algorithmic biases, all of which are central to ethical and effective AI implementation in mental health settings.

Further research should explore integrating broader psychological constructs, such as emotion regulation, cognitive biases, and attachment theory, alongside additional case formulation models, including standard CBT or psychodynamic approaches, to inform future AI development. This could enhance ChatGPT's ability to generate contextually accurate feedback, particularly for emotionally complex scenarios. Additionally, longitudinal studies assessing the AI's performance across diverse clinical populations, cultural contexts, and diagnostic categories are necessary to determine its long-term utility and reliability as a support tool for clinicians.

Conclusions

This study explored the potential of GPT-based algorithms, such as ChatGPT, to support psychological analysis through structured reasoning aligned with a theoretical framework. Findings indicate that the model can produce coherent feedback using LIBET-related terminology and structure, which may assist reflective processes, like in early-stage case formulation. However, its limited capacity for abstraction and emotional contextualization underscores the need for refinement.

Future research should prioritize improving the interpretative depth and emotional nuance of AI-generated outputs. Collaborative efforts between clinicians and AI developers will be essential to ensure that such tools complement—rather than replace—clinical judgment. While integration into therapeutic settings remains premature, these technologies may hold value as assistive instruments in structured reasoning tasks when used with appropriate safeguards and critical oversight.

Conflict of Interest

The authors have no conflicts of interest to declare.

Use of AI Services

The authors declare they have used AI services, specifically ChatGPT (OpenAI, GPT-4 and GPT-4o), both as the object of study (for qualitative analysis of its feedback in psychological case conceptualization) and as an assistive tool for grammar correction and style refinement. AI assistance was strictly limited to improving readability, grammar, and coherence. It did not contribute to conceptualization, data analysis, or scientific reasoning. All intellectual content, theoretical interpretations, and conclusions are the sole work of the authors.

Authors’ Contribution

Matilde Buattini: conceptualization, investigation, methodology, visualization, writing—original draft, writing—review & editing, project administration. **Donald Barjami:** conceptualization, methodology, writing—original draft. **Lorenza Paponetti:** investigation, methodology, visualization. **Dalila Torres:** visualization, writing—review & editing. **Rosita Borlimi:** supervision, validation. **Gabriele Caselli:** supervision, validation.

Acknowledgement

The authors extend their gratitude to the participants who generously contributed their time and insights to this study. We also wish to thank both the Lab MeThe and the LIBET Lab at Sigmund Freud University for their invaluable support and resources throughout the research process.

Data Availability Statement

Data are available on reasonable request by contacting the corresponding author.

Funding Statement

The author received no financial support for the research, authorship, and/or publication of this article.

Ethics Approval

This study received ethical approval from the Ethics Committee of Sigmund Freud University (Reference: PD48HK50C37GJ490824).

References

Bambling, M., King, R., Raue, P., Schweitzer, R., & Lambert, W. (2006). Clinical supervision: Its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research, 16*(3), 317–331. https://doi.org/10.1080/10503300500268524

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond* (2nd ed.). The Guilford Press.

Beckes, L., IJzerman, H., & Tops, M. (2015). Toward a radically embodied neuroscience of attachment and relationships. *Frontiers in Human Neuroscience, 9*, Article 266. https://doi.org/10.3389/fnhum.2015.00266

Beg, M. J. (2025). Responsible AI integration in mental health research: Issues, guidelines, and best practices. *Indian Journal of Psychological Medicine, 47*(1), 5–8. https://doi.org/10.1177/02537176241302898

Beg, M. J., & Verma, M. K. (2024). Exploring the potential and challenges of digital and AI-driven psychotherapy for ADHD, OCD, Schizophrenia, and substance use disorders: A comprehensive narrative review. *Indian Journal of Psychological Medicine.* Advance online publication. https://doi.org/10.1177/02537176241300569

Beg, M. J., Verma, M., Vishvak Chanthar, K. M. M., & Verma, M. K. (2024). Artificial intelligence for psychotherapy: A review of the current state and future directions. *Indian Journal of Psychological Medicine, 47*(4), 314–325. https://doi.org/10.1177/02537176241260819

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D.,… Liang, P. (2021). *On the opportunities and risks of foundation models*. arXiv. https://doi.org/10.48550/arXiv.2108.07258

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021a). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health, 13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Braun, V., & Clarke, V. (2021b). *Thematic analysis: A practical guide*. SAGE Publications.

Brown, J. R., Holloway, E. D., Akakpo, T. F., & Aalsma, M. C. (2014). “Straight up”: Enhancing rapport and therapeutic alliance with previously-detained youth in the delivery of mental health services. *Community Mental Health Journal, 50*(2), 193–203. https://doi.org/10.1007/s10597-013-9617-3

de Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review, 85*, Article 102002. https://doi.org/10.1016/j.cpr.2021.102002

de Jong, K., Douglas, S., Wolpert, M., Delgadillo, J., Aas, B., Bovendeerd, B., Carlier, I., Compare, A., Edbrooke-Childs, J., Janse, P., Lutz, W., Moltu, C., Nordberg, S., Poulsen, S., Rubel, J. A., Schiepek, G., Schilling, V. N. L. S., van Sonsbeek, M., & Barkham, M. (2025). Using progress feedback to enhance treatment outcomes: A narrative review. *Administration and Policy in Mental Health and Mental Health Services Research, 52*(1), 210–222. https://doi.org/10.1007/s10488-024-01381-3

De Paoli, S., & Mathis, W. S. (2025). Reflections on inductive thematic saturation as a potential metric for measuring the validity of an inductive thematic analysis with LLMs. *Quality & Quantity, 59*(1), 683–709. https://doi.org/10.1007/s11135-024-01950-6

Enlow, P. T., McWhorter, L. G., Genuario, K., & Davis, A. (2019). Supervisor–supervisee interactions: The importance of the supervisory working alliance. *Training and Education in Professional Psychology, 13*(3), 206–211. https://doi.org/10.1037/tep0000243

John, S., & Segal, D. L. (2015). Case conceptualization. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology*, (pp. 1–4). John Wiley & Sons. https://doi.org/10.1002/9781118625392.wbecp106

Jones, C., & Bergen, B. (2023). *Does GPT-4 Pass the Turing Test?* arXiv. https://doi.org/10.48550/arXiv.2310.20216

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Kjell, O. N. E., Kjell, K., & Schwartz, H. A. (2024). Beyond rating scales: With targeted evaluation, large language models are poised for psychological assessment. *Psychiatry Research, 333*, Article 115667. https://doi.org/10.1016/j.psychres.2023.115667

Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *The New England Journal of Medicine, 388*(13), 1233–1239. https://doi.org/10.1056/NEJMsr2214184

Liu, J. (2024). ChatGPT: Perspectives from human–computer interaction and psychology. *Frontiers in Artificial Intelligence, 7*, Article 1418869. https://doi.org/10.3389/frai.2024.1418869

McCoy, R. T., Yao, S., Friedman, D., Hardy, M. D., & Griffiths, T. L. (2024). Embers of autoregression show how large language models are shaped by the problem they are trained to solve. In *Proceedings of the National Academy of Sciences of the United States of America, 121*(41), Article e2322420121. https://doi.org/10.1073/pnas.2322420121

Norman, K. P., Govindjee, A., Norman, S. R., Godoy, M., Cerrone, K. L., Kieschnick, D. W., & Kassler, W. (2020). Natural language processing tools for assessing progress and outcome of two veteran populations: Cohort study from a novel online intervention for posttraumatic growth. *JMIR Formative Research, 4*(9), Article e17424. https://doi.org/10.2196/17424

Offredi, A., Bertozzi, G., Carbonari, A., Dellasanta, F., Galanti, A., & Caselli, G. (2025). Comparison between clinical and non-clinical populations using the Life and Themes Implication Biased: Elicitation and Treatment Questionnaire (LIBET-Q): A contribution to the validation of the LIBET model. *Psicoterapia Cognitiva e Comportamentale, 31*(1), 39–60. https://doi.org/10.14605/PCC3112502

OpenAI. (2024). *OpenAI.* Retrieved (December 15, 2023), from https://openai.com/

OpenAI Platform. (n.d.). Prompt engineering. OpenAI Platform. Retrieved (December 15, 2023), from https://platform.openai.com/docs/guides/prompt-engineering

Pham, K. T., Nabizadeh, A., & Selek, S. (2022). Artificial intelligence and chatbots in psychiatry. *The Psychiatric Quarterly, 93*(1), 249–253. https://doi.org/10.1007/s11126-022-09973-8

Sassaroli, S., Caselli, G., Mansueto, G., Palmieri, S., Pepe, A., Veronese, G., & Ruggiero, G. M. (2022). Validating the diathesis–stress model based case conceptualization procedure in cognitive behavioral therapies: The LIBET (Life Themes and Semi-Adaptive Plans— Implications of Biased Beliefs, Elicitation and Treatment) procedure*. Journal of Rational-Emotive & Cognitive-Behavior Therapy, 40*(3), 527–565. https://doi.org/10.1007/s10942-021-00421-3

Sassaroli, S., Ruggiero, G. M., & Caselli, G. (2023). *Capire il paziente. Guida alla formulazione del caso LIBET: Dalla teoria all'applicazione* [Understanding the Patient. A Guide to LIBET Case Formulation: From Theory to Practice]. Giunti Psicologia.

Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R. J., Willer, R., & Eichstaedt, J. C. (2024). Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. *NPJ Mental Health Research, 3*, Article 12. https://doi.org/10.1038/s44184-024-00056-z

Tanana, M. J., Soma, C. S., Kuo, P. B., Bertagnolli, N. M., Dembe, A., Pace, B. T., Srikumar, V., Atkins, D. C., & Imel, Z. E. (2021). How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behavior Research Methods, 53*(5), 2069–2082. https://doi.org/10.3758/s13428-020-01531-z

Van Le, D., Montgomery, J., Kirkby, K. C., & Scanlan, J. (2018). Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. *Journal of Biomedical Informatics, 86*, 49–58. https://doi.org/10.1016/j.jbi.2018.08.007

Weck, F., Jakob, M., Neng, J. M. B., Höfling, V., Grikscheit, F., & Bohus, M. (2016). The effects of bug-in-the-eye supervision on therapeutic alliance and therapist competence in cognitive-behavioural therapy: A randomized controlled trial. *Clinical Psychology & Psychotherapy, 23*(5), 386–396. https://doi.org/10.1002/cpp.1968

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology, 5*, Article 506. https://doi.org/10.3389/fpsyg.2014.00506

Wendler, C., Veselovsky, V., Monea, G., & West, R. (2024). *Do llamas work in English? On the latent language of multilingual transformers*. arXiv. https://doi.org/10.48550/arXiv.2402.10588

Wong, R. S.-Y. (2024). ChatGPT in psychiatry: Promises and pitfalls. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery, 60*(1), Article 14. https://doi.org/10.1186/s41983-024-00791-2

Appendices

Appendix A

Custom GPT Configuration Log

Disclaimer: The following exchange reflects the configuration process of the custom MyGPT developed for the study. For consistency with the manuscript, the interaction is presented in English, while the chatbot itself operates in Italian.

ChatGPT - GPT Builder. Hi! I'll help you build a new GPT. You can say something like, "make a creative who helps generate visuals for new products" or "make a software engineer who helps format my code." What would you like to make?

Researcher. I want to create a chatbot in Italian that knows the theory of the LIBET model (a model for understanding individual functioning, explained in the attached PDF), specifically life themes and semi-adaptive plans. To this end, learn the concepts contained in the PDF.

The chatbot will be used to upload interview transcripts conducted with non-clinical participants and will be asked to provide feedback based on the aforementioned theoretical framework.

[LIBET Manual.pdf attached; Sassaroli et al., 2024]

ChatGPT - GPT Builder. I have acquired the information about the LIBET model from the document you provided. The LIBET model (Life themes and semi-adaptive plans: Implications of biased Beliefs, Elicitation, and Treatment) is a psychotherapeutic approach that integrates cognition and the patient's personal and developmental history. It focuses on identifying life themes and semi-adaptive plans, which influence the patient's dysfunctional behavior, and adopts a personalized and collaborative therapeutic approach to treatment. Now that I know the details, let's proceed to configure your GPT. How would you like this GPT to behave and respond?

Researcher. The chatbot will be used to upload transcripts of interviews conducted with non-clinical participants, and it will be asked to provide feedback based on the theoretical framework just learned. It must speak in Italian, maintain a formal tone, and use psychological terminology consistent with the theoretical framework. Always remember that the bot is speaking with "I" (the Interviewer or psychologist) about "P" (the Participant or interviewee), so it must use the third person.

ChatGPT - GPT Builder. I've configured the GPT as a LIBET Model Analyst. It's ready to analyze interview transcripts, using the terminology and concepts of the LIBET model in Italian, and responding formally. Try asking it something! You can start with phrases like "Could you analyze this transcript?" or "What is your feedback on this interview?" If you need to make further adjustments, let me know!

Researcher. Before testing, some instructions are added: It is necessary to emphasize the "reasoning" behind each response. Avoid inserting information that is not requested or not relevant to the theoretical framework. Never use bullet points, but respond with a text of the required length stated in the prompt.

ChatGPT - GPT Builder. I've updated the instructions for the LIBET Model Analyst to emphasize the reasoning behind each response, avoid irrelevant information, and avoid using bullet points, respecting the required length specified in the prompt. Would you like to try the GPT now or is there anything else you'd like to modify?

Researcher. The Title: instead of "LIBET Model Analyst" I want to call it "LIBET-Chat Interviste"

ChatGPT - GPT Builder. I've changed the name to LIBET-Chat Interviews. Are you ready to try this GPT or is there anything else you'd like to adjust before getting started?

Researcher. [no answer]

Appendix B

Prompt in Italian (Original Version) for Life-Themes ChatGPT's Analysis

Alla luce del trascritto di un'intervista (propriamente pulita da dati sensibili) svolta con un partecipante non-clinico (P) e della teoria di riferimento sul tema di vita: arriva ad ipotizzare il tema di vita dell'intervistato (P) tra i 3 temi di vita possibili, facendo prima emergere il ragionamento e le motivazioni che ti portano a fare questa ipotesi (scrivi 200-300 parole in tutto).

Di seguito la teoria:

***

Il tema di vita è una sensibilità fisiologica, emotiva e cognitiva ai segnali problematici esterni e interni. Tale sensibilità o vulnerabilità si apprende nel processo di crescita e può divenire eccessiva se l'individuo polarizza l'attenzione su una particolare frustrazione di un determinato bisogno.

1. Tema di Minaccia e Inadeguatezza: sensazione di pericolo, carenza di sicurezza personale e protezione, non garantita materialmente e affettivamente dalle figure significative e affidabili che dovrebbero fornire nell'età evolutiva protezione fisica, nutrimento e accudimento. Questo si può anche legare a un non sentirsi all'altezza delle prestazioni richieste dall'ambiente, sia in termini di capacità fisiche sia di competenze sociali e relazionali.
2. Tema Disamore e Inadeguatezza: sensazione di perdita di senso, inutilità, esclusione, assenza di valore, collegato a stati emotivi di tristezza e depressione. Gli stati depressivi sono legati a un'atmosfera nel periodo di crescita di deprivazione emotiva, affettività fredda e distanziante, che disconosce l'affettività e in cui i contatti corporei non sono teneri, ma rari e impacciati.
3. Tema di Indegnità e Inadeguatezza: sensazione di inferiorità, tossicità, disprezzo verso di sé, umiliazione e vergogna, che può originare in famiglie in cui è vigente uno stile relazionale criticante, normativo, controllante, esigente e oppressivo, in cui i valori regolativi sono vissuti e trasmessi in maniera colpevolizzante e punitiva e le valutazioni negative sono associate alle forme di autoaffermazione o autorealizzazione non conformi a questi valori.

***

Di seguito il trascritto. In [] le info eliminate per motivi di privacy o aspetti non verbali.

[trascritto]

Prompt Translated in English for Life-Themes ChatGPT's Analysis

In light of the transcript of an interview (properly cleansed of sensitive data) conducted with a non-clinical participant (P) and the theoretical framework on the life theme: hypothesize the life theme of the interviewee (P) among the 3 possible life themes, first outlining the reasoning and motivations that lead you to make this hypothesis (write 200-300 words in total).

Below is the theory:

***

The life theme is a physiological, emotional, and cognitive sensitivity to external and internal problematic signals. This sensitivity or vulnerability is learned during the growth process and can become excessive if the individual polarizes attention on a particular frustration of a specific need.

1. Threatened and Inadequate Theme: A feeling of danger, lack of personal safety and protection, not guaranteed materially or emotionally by significant and reliable figures who, during the developmental age, should provide physical protection, nourishment, and care. This can also relate to a sense of not feeling up to the demands of the environment, both in terms of physical abilities and social and relational skills.
2. Unloved and Inadequate Theme: A feeling of loss of meaning, uselessness, exclusion, absence of value, linked to emotional states of sadness and depression. Depressive states are tied to an atmosphere

during the growth period characterized by emotional deprivation, cold and distancing affectivity, which disregards emotional connection, and where physical contact is not tender but rare and awkward.

3. Unworthy and Inadequate Theme: A feeling of inferiority, toxicity, self-contempt, humiliation, and shame, which may originate in families with a critical, normative, controlling, demanding, and oppressive relational style, where regulatory values are experienced and conveyed in a guilt-inducing and punitive manner, and negative evaluations are associated with forms of self-assertion or self-realization that do not conform to these values.

Below is the transcript. In [] are the omitted details for privacy or non-verbal aspects.

[transcript]

About Authors

Matilde Buattini is a clinical psychologist and sexologist, currently a Ph.D. candidate at Sigmund Freud University in Vienna, conducting research on AI's applications in psychotherapy. She holds a Master's degree in Clinical Psychology from Sigmund Freud University, Milan, and is specializing in Cognitive-Behavioral Psychotherapy.

https://orcid.org/0000-0001-6219-6457

Donald Barjami holds a Bachelor's degree in Cognitive and Psychobiological Psychological Sciences from the University of Padua and a Master's degree in Psychology from Sigmund Freud University, Milan. He completed a post-graduate clinical internship at Il Porto Onlus.

Lorenza Paponetti is a Clinical Psychologist, specializing in Cognitive-Behavioral Psychotherapy. She holds a Master's degree in Psychology from Sigmund Freud University, Milan.

Dalila Torres is a clinical psychologist and Ph.D. candidate at Sigmund Freud University in Wien, specializing in forensic psychology and metacognition. She collaborates with correctional institutions and forensic psychiatry units, conducting research on psychopathological profiles and legal processes.

https://orcid.org/0009-0003-7728-243X

Rosita Borlimi holds a Ph.D. in Psychotherapy and specialized in Health Psychology from the University of Bologna. She is a Lecturer and the Director of the Affective Neuroscience Lab at Sigmund Freud University, Milan, and works as a psychotherapist in private practice and at the Psychological Support Service (SAP) of the University of Bologna.

https://orcid.org/0000-0001-5382-0545

Gabriele Caselli is Full Professor in Clinical Psychology and Scientific Director of Studi Cognitivi, Milan. He holds a Ph.D. in Clinical Psychology and specializes in Cognitive-Behavioral and Metacognitive Therapy.

https://orcid.org/0000-0002-5159-7164

✉ **Correspondence to**
Matilde Buattini, Sigmund Freud University Milano, MeThe Research Lab, Psychology Department, Ripa di Porta Ticinese 77, 20143, Milan, Italy; Sigmund Freud PrivatUniversität, Psychology Department, Freudplatz 1, A-1020, Wien, Austria; Studi Cognitivi, Foro Buonaparte 57, 20121, Milan, Italy, buattini.phd@milano-sfu.it

© Author(s). The articles in Cyberpsychology: Journal of Psychosocial Research on Cyberspace are open access articles licensed under the terms of the Creative Commons BY-SA 4.0 International License which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited and that any derivatives are shared under the same license.