Article

Towards Integrated AI Psychotherapy Supervision: A Proposal for a ChatGPT-4 Study

Valeria Cioffi *, Ottavio Ragozzino , Chiara Scognamiglio , Lucia Luciana Mosca , Enrico Moretto , Roberta Stanzione , Francesco Marino , Annamaria Acocella , Angela Ammendola , Rossella D'aquino , Simona Durante , Enrica Tortora , Flavia Morfini , Caludia Montanari , Veronica Rosa , Oliviero Rossi , Antonio Ferrara , Elisa Mori , Elena Gigante , Mariano Pizzimenti , Sonia Zangarini , Raffaele Sperandeo , Daniela Cantone

Posted Date: 13 January 2025

doi: 10.20944/preprints202501.0848.v1

Keywords: Psychotherapy Supervision; Artificial Intelligence (AI); Large Language Models (LLMs); ChatGPT; Gestalt Therapy (GT)

Article

Towards Integrated AI Psychotherapy Supervision: A Proposal for a ChatGPT-4 Study

Valeria Cioffi [1,*], Ottavio Ragozzino [1], Chiara Scognamiglio [1], Lucia Luciana Mosca [1], Enrico Moretto [1], Roberta Stanzione [1], Francesco Marino [1], Annamaria Acocella [2], Angela Ammendola [1], Rossella D'Aquino [1], Simona Durante [1], Enrica Tortora [1], Flavia Morfini [1], Claudia Montanari [3], Veronica Rosa [3], Oliviero Rossi [2], Antonio Ferrara [4], Elisa Mori [5], Elena Gigante [6], Mariano Pizzimenti [7], Sonia Zangarini [4], Raffaele Sperandeo [1] and Daniela Cantone [8]

[1] SiPGI–Postgraduate School of Integrated Gestalt Psychotherapy, Torre Annunziata, Italy
[2] IPGE Istituto di Psicoterapia della Gestalt Espressiva - Via Costantino Morin, 24 - 00195 Roma
[3] ASPIC Scuola di Psicoterapia Via Vittore Carpaccio, 32 - 00147 Roma
[4] iGAT Istituto di Psicoterapia della Gestalt e Analisi Transazionale - Via Pirro Ligorio, 20 80129 Napoli
[5] IGA Istituto Gestalt Analitica - Via Padre Semeria 33 - 00154 Roma
[6] SIPGI Scuola in Psicoterapia gestaltica integrata - Via Abruzzo, 6, 91100 Trapani (TP)
[7] SGT Scuola Gestalt Torino - Via Po 14 10123 Torino
[8] Università degli Studi della Campania - Luigi Vanvitelli
* Correspondence: valeria.cioffi@gmail.com

Abstract: Psychotherapeutic supervision is crucial for the quality of the development of professionals and the promotion of well-being and therapeutic effectiveness of trainees. Emerging computational technologies can offer new potentials to supervision practice. The aim of this paper is to test the current capabilities of the GPT-4 linguistic model to support the work of psychotherapists, providing secondary and complementary feedback to supervision. This paper represents the initial phase (phase 0) useful for laying the foundations and adequately calibrating the methodology of a future more extensive work which will consist of further and more detailed training of GPT-4 on the tasks required. The availability of this possibility, for a trainee or a therapist, would mean significant support for self-monitoring and increasing this competence, an incentive and support for awareness of the need to resort to a supervision meeting. This could result in enhanced training, improved efficiency and clinical effectiveness, maintenance of well-being, prevention of burnout and improved patient outcomes.

Keywords: psychotherapy supervision; artificial intelligence (AI); large language models (LLMs); ChatGPT; gestalt therapy (GT)

1. Introduction

In psychotherapeutic practice, supervision is a functional element for the quality of the development of professionals, the promotion of their well-being and therapeutic effectiveness [1]. Furthermore, it is crucial in the orientation of trainees and in increasing their therapeutic skills in a safe and supportive context. In fact, through supervision, trainees are led to explore and implement their therapeutic skills according to interactive and experiential methods. [2]. It is recognized by supervision a facilitator role for the acquisition of therapeutic skills, for the management of the emotional depletion that can derive from the exercise of clinical practice [3] Rønnestad, M. H. et al. (2019) [3], for example, highlight how supervision represents a key factor in managing the emotional impoverishment that can arise from clinical practice. In particular, they emphasize how it acts as an essential catalyst for the professional development of therapists, as well as for the purposes of stress management, above all, for those who are at the beginning of their careers, for example postgraduates. The supervisory relationship, in fact, implies emotional and professional support with a mitigating effect with respect to professional stressors, configuring itself as a functional element for the prevention of burnout [4]. The well-being of the therapist is an indispensable condition for his efficiency and effectiveness, it is evident that the quality of the therapeutic intervention [5].

Technological developments offer new potentials to clinical practice [6–9]. In fact, supervision can benefit from the potential and ease of use of Artificial Intelligence. Artificial intelligence systems of the "Generative Pre- Trained Transformer" type (GPT, generative pre-trained transformer), are linguistic models that, through Deep Learning, learn to produce language similar to human language and to provide correct responses to inputs. In the latest version, GPT-4, it has been implemented the possibility of reinforcement learning using human feedback [10].

2. Aim

The aim of this paper is to test the current capabilities of the GPT-4 linguistic model in supporting the work of psychotherapists by providing secondary and complementary feedback during supervision.

Specifically, we intend to test to what degree the model is able to identify, starting from a specific prompt provided by the therapist, specific elements relating to the session in question. This work is considered as the initial phase (phase 0) useful for laying the foundations and adequately calibrating the methodology of a future more extensive work which will consist of further and more detailed training of GPT-4 on the tasks required.

3. Methodology

The methodology used for this pilot study consists of 5 macro-phases:

1) Verification of the linguistic model's skills in generating text that shows knowledge of Gestalt psychotherapy and supervision. This preliminary verification also involved testing the linguistic model's ability to generate transcripts of imaginary psychotherapy and supervision sessions, consistent with the professional and personal information provided on the different imaginary characters and which had distinctive characteristics typical of Gestalt Therapy (GT).
2) Verification of ChatGPT 4's ability to provide feedback to support the supervision of a real case in the context of a chat window in which the topic had already been explored in depth.
3) Testing ChatGPT 4's ability to provide feedback to support the supervision of the same real case in a clean conversation window.
4) Creation of a prompt that acts as an incipit to pre-direct the model to deal in more depth with the subject of interest.
5) Test of the linguistic model's ability to provide adequate feedback, after entering the aforementioned prompt, providing information on the same case, patient and therapist in different ways.

4. Results

The verification of ChatGPT's knowledge on GT showed excellent results even when trying to go into various details and questioning the bot on subtleties.

This verification also involved asking the bot to generate verbatim of psychotherapy sessions: for a first session, very general information about the clinical case was provided to the bot and for a second session, more detailed information was provided on both the patient and the therapist, including both professionally and personally.

The two verbatim generated, in addition to taking on realistic characteristics both from the point of view of verbal interaction and regardin the non-verbal and paraverbal aspects described by GPT-4, show specific characteristics such as:

Here-and-now: the therapist (called Paolo) invites the patient (called Vito) to focus on the "here-and-now": on the current sensations, emotions and thoughts that emerge during the session. This is a central aspect of GT, where the emphasis is placed on what is immediately present, rather than on past or hypothetical future events.

Emotional and bodily awareness: Paolo leads Vito to focus on the bodily sensations linked to his feelings of abandonment. This encourages the patient to explore and accept their physical reactions as an integral part of their emotional experience. This is also a characterizing aspect of GT.

Enhancement of self-regulation: Paolo encourages Vito to use breathing as a tool to independently manage pain and stress. The Gestalt approach indeed emphasizes self- regulation, helping the client develop internal skills to deal with difficult situations.

Process of acceptance and responsibility: The patient is guided towards acceptance that he does not have to be perfect and the recognition that he can take a break. This is fundamental in Gestalt, where self-recognition and acceptance are seen as steps towards healing.

After the imaginary session described in the second verbatim, the chatbot was asked to imagine that the therapist wrote his reflections in a diary, and then asked a more experienced colleague for supervisory advice. The linguistic model was then provided to imagine a new character (called Valeria), the supervisor, providing her personal and professional characteristics and related to the knowledge of the supervised therapist. Next, ChatGPT-4 was asked to generate a verbatim of this imaginary supervision. In this case a small inconsistency was observed for which it was necessary to provide feedback requesting a correction.

After supervision, the request provided to the chatbot was to imagine that the therapist wrote down his reflections again, and then carried out a further psychotherapy session with the same patient. Analyzing this last verbatim, it emerged that supervision played a fundamental role in improving the therapeutic approach. Here, Paolo applies breathing and self-exploration techniques with greater confidence and method, the result of the reflections and feedback integrated during supervision. He has taken a more structured approach, using specific exercises that prepare patients to handle stressful situations. This evolution also shows Paolo's greater awareness and management of emotional distance, allowing him to guide patients with greater objectivity and calm. Furthermore, his ability to openly discuss the used techniques increases patients' trust, strengthening the therapeutic relationship.

Moving at this point from the imaginary to the real, continuing to operate in the same conversation window, ChatGPT was provided with an in-depth description, appropriately anonymized, of a real clinical case (which we will call "Leonard Case") followed by one of the authors of this study: a therapy process at its beginning, with the description of the first 3 sessions, the previous psychodiagnostics analysis (conducted by the same professional) and the related setting transition.

In order to request supervisory feedback from the chatbot, detailed information about the therapist was provided to it, asking the bot itself what information would be useful to respond to our request.

The question asked to ChatGPT-4 was to obtain feedback from both Valeria (previously described supervisor-character) and Paolo (the aforementioned therapist- character). Before giving us the answer, the chatbot specifically asked what aspect we were requesting supervision on: use of Gestalt techniques, management of the therapeutic relationship or interpretation of Leonard's personality functioning.

In response, the request asked was to provide feedback both on the techniques and on the management of the relationship: in both cases we obtained valid and useful responses in the judgment of the therapist-coauthor in question and, each feedback, consistent with the characteristics of the two characters who took on the role of supervisor. Subsequently, the same detailed formulation of the Leonard case was proposed again to ChatGPT, but this time in a new conversation window, starting from scratch. Supervisory feedback was then requested regarding the effectiveness of the techniques used by the therapist. In this case, the feedback obtained was scant and excessively generic. The substantial difference is attributable to the fact that in the first window aspects relating to GT had been carefully explored, while in the second the request had been made "cold". For this reason, the hypothesis was that, to obtain accurate and reliable feedback, it was necessary to develop a prompt that would act as an initial, theoretical and background framework for the topics covered. For this purpose, by proposing as input material the same information on the Gestalt provided to us by ChatGPT during the initial "query", the chatbot itself was asked to develop a prompt that could pre-orientate and direct us towards the knowledge niche in which we was interested in operating.

Finally, we proceeded with the test of the actual usefulness of the prompt developed by ChatGPT, operating each time in new conversation windows, each time providing first the developed prompt-frame and then a different formulation of the Leonard case, each time with different focuses and different ways, forms and levels of detail in providing information about the case, the therapist and the sessions. Specifically, the information on the case was provided in the previously initial form of the complete formulation, description of the psychodiagnostics assessment, psychodiagnosis according to DSM-5 [11], the therapist's reflections on the transition of setting from psychodiagnostics to psychotherapeutic and description of the first 3 sessions; the description of the case was also proposed to the chatbot through the description of only the first session or a fragment of it, in both cases anticipated either by an incipit relating to the essential anamnestic and diagnostic information, or by a more in-depth presentation. As regards information about the therapist, it was provided to the chatbot both through a free prosaic description and in a more concise form providing targeted information. In providing information about the therapist, a self-administered questionnaire was also used to explore the interpersonal process in the relationship with a client [1].

In each attempt, reliable, potentially useful and adequately detailed feedback on the therapeutic work was obtained. The greatest level of detail and in-depth analysis was found in the first conversation window, the same in which there had previously been a long and in-depth dialogue on the matter of interest. All the outputs obtained by providing the information to ChatGPT 4 after entering the specifically generated prompt showed reasonable coherence, accuracy, and were all considered quite useful by the therapist in question. The output obtained in the tests for which the questionnaire for the exploration of the interpersonal process was also provided placed the focus more on the therapist's experience.

The knowledge achieved through this pilot study lays the foundations for subsequent in-depth phases.

5. Conclusions

We observed that by adequately pre-addressing the chatbot on the area of interest, more adequate and useful responses for the purpose were obtained. The accuracy and validity of these answers was found both in the first experiment, in which the case study was entered into the same conversation window in which there had previously been an in-depth interaction regarding GT and supervision, and in the different conversation windows in which the prompt generated by ChatGPT 4 based on its previously provided answers was used as an incipit. The validity of these answers remained unchanged regardless of the ways in which information on the case, the therapist and the patient were provided. Only in the tests in which ChatGPT 4 was also provided with the compilation of a self-report by the therapist regarding his experiences during the psychotherapy session brought under supervision, did the focus of the response provided by ChatGPT shift more to the experiences of the therapist. The aim is described as the evaluation of the current capabilities of ChatGPT-4 to provide therapists with useful feedback, and the best ways of interacting with the chatbot in order to obtain it, which can integrate supervision, tools and classic support contexts to clinical practice, therapist training, and motivational incentive. The availability of this possibility, for a therapist, would mean significant support for self-monitoring and increasing this competence, an incentive and support for awareness of the need to resort to a supervision meeting. This could result in enhanced training, improved efficiency and clinical effectiveness, maintenance of well-being, prevention of burnout and improved patient outcomes.

6. Future Developments

The present stands as "phase 0", preliminary to a broader and long-term study. The details and characteristics of this study will be calibrated in more detail on the basis of the outcomes of this phase: 1) Reports of psychotherapy sessions and key information relating to the therapists who conducted them will be collected and used to better experiment and refine the methods of interaction with ChatGPT here to experiment in order to have more accurate feedback.

2) Knowledge of the default capabilities of ChatGPT-4.0 to assist us in the above contexts, will be the starting point for structuring a further training program of the linguistic model, through a fine-tuning process [12], in which the reports of the sessions of psychotherapy will be used as input and those of real supervision sessions relating to them will be paired as desired outputs for training purposes. The resulting refinement of skills could then be exposed to continuous feedback loops provided by the therapists themselves.

7. Limitations

The development of AI-based technologies is increasingly rapid and unpredictable. The basic characteristics of the linguistic model considered could change significantly in a short time. Although it is safe to assume that these changes go in the direction of improvement, this could still influence what might be, in the future, the most effective ways of interacting with the machine to achieve a given result.

Considering also that the study to come, of which the one proposed in this paper is the basis, would take longer, we believe it is equally risky to predict how these developments will evolve in the meantime.

Therefore, the risk is highlighted that the results of this work could prove no longer suitable for adequate planning of the future training protocol. As the "starting point" it could be different just as the operating methods of the technologies in question may prove to be different from the current ones. While the results of the fine-tuning training envisaged for the future evolution of the present study could, at the conclusion of the same, prove to be even obsolete as they are negligible and superfluous compared to what could be the basic capabilities of intelligence in the near future artificial.

Funding: This study did not receive external funding.

Conflicts of Interest: The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Giusti, E., Montanari, C., Spalletta, E. (2000). *La supervisione clinica integrata. Manuale di formazione pluralistica in counseling e psicoterapia*. Italia: Elsevier.
2. Yontef, G. (1996). Supervision from a Gestalt therapy perspective. *British Gestalt Journal*, 5, 92-102.
3. Rønnestad, M. H., Orlinsky, D. E., Schröder, T. A., Skovholt, T. M., &amp; Willutzki, U. (2019). The professional development of counsellors and psychotherapists: Implications of empirical studies for supervision, training and practice. *Counselling and Psychotherapy Research, 19(3)*, 214-230.
4. Schavel, M., Kuzysin, B., Beresova, A., &amp; Hunyadiová, S. (2018). The impact of supervision in social work on the burnout syndrome prevention. *Przestrzeń Społeczna, 2(2*/2018 (16).
5. Watkins Jr, C. E. (2011). Does psychotherapy supervision contribute to patient outcomes? Considering thirty years of research. *The clinical supervisor, 30(2)*, 235- 256.
6. Cioffi, V., Mosca, L. L., Moretto, E., Ragozzino, O., Stanzione, R., Bottone, M., ... & Sperandeo, R. (2022). Computational Methods in Psychotherapy: A Scoping Review. *International Journal of Environmental Research and Public Health, 19(19)*, 12358.
7. Cantone, D., Guerriera, C., Architravo, M., Alfano, Y. M., Cioffi, V., Moretto, E., ... & Sperandeo, R. (2021). A sample of Italian psychotherapists express their perception and opinions of online psychotherapy during the covid-19 pandemic. *Rivista di psichiatria*, *56*(4), 198-204.

8. Cioffi, V., Cantone, D., Guerriera, C., Architravo, M., Mosca, L. L., Sperandeo, R., ... & Maldonato, N. M. (2020, September). Satisfaction degree in the using of VideoConferencing Psychotherapy in a sample of Italian psychotherapists during Covid-19 emergency. In *2020 11th IEEE international conference on cognitive Infocommunications (CogInfoCom)* (pp. 000125-000132). IEEE.
9. Sperandeo, R., Cioffi, V., Mosca, L. L., Longobardi, T., Moretto, E., Alfano, Y. M., ... & Maldonato, N. M. (2021). Exploring the question:"Does empathy work in the same way in online and in-person therapeutic settings?". *Frontiers in psychology*, *12*, 671790.
10. Kalyan, K. S. (2023). A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal*, 100048.
11. American Psychiatric Association. (2014). *Manuale diagnostico e statistico dei disturbi mentali* (5ª ed.). Raffaello Cortina Editore.
12. Rothman, D. (2024). *Transformers for Natural Language Processing and Computer Vision*. Packt Publishing.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.