Tacit Psychoanalytic Knowledge and the Limits of Artificial Intelligence: A Case Report

Husna Manalai[1,2] ; Partam Manalai, MD [2,3]; Priya Batheja[4] , DO (2027); BA (2026); Gul PM Osmani, MD[2]; Suneeta Kumari[5] , MD; Mansoor Malik[6] , MD

1.Bard College, 30 Campus Rd, Annandale-on-Hudson, NY 12504 ; 2.Intuitive Insight, Inc, 1765 Greensboro Station Pl, Suite 900, McLean, VA 22102; 3.Mary Washington Healthcare, 1200 Sam Perry Boulevard, Fredericksburg, VA 22401; 4.Edward Via College of Osteopathic Medicine – Virginia Campus; 2265 Kraft Drive;Blacksburg, VA 24060; 5.Sinai Hospital of Baltimore, 2401 Belvedere Avenue, Baltimore, MD 21215; 6.Johns Hopkins University, 4940 Eastern Avenue, Alpha Commons 4 th Floor, Baltimore, MD 21224

ABSTRACT:

Artificial intelligence (AI) is rapidly transforming many areas of medicine, yet its relevance to expressive and psychoanalytic psychotherapies remains limited. Unlike protocol-driven interventions, psychoanalytic treatment depends on tacit, experiential knowledge cultivated and transmitted over years of training and supervision. Owing to ethical, conceptual, and epistemological concerns, experienced psychoanalytically trained clinicians are often less inclined to engage with AI, thus, core psychoanalytic capacities risk being diluted or lost as experienced clinicians retire.

This article presents a case report centered on a supervisory intervention unlikely to be generated by current AI systems. A 50-year-old man with a history of childhood and early adult sexual abuse developed complex grief after the death of his abuser while in insight-oriented psychotherapy with a PGY-4 psychiatry resident. In supervision, the psychoanalyst proposed an ethically challenging yet clinically precise inquiry into the patient's possible experience of sexual sensation during the abuse. Though counterintuitive, the intervention deepened exploration, reduced pathological guilt, and supported emotional integration.

We examine whether contemporary large language models could generate such an intervention. We argue that without active psychoanalytic engagement and deliberate articulation of contextualized clinical reasoning, AI will remain limited as psychoanalysts' participation is essential if AI is to become clinically useful.

INTRODUCTION:

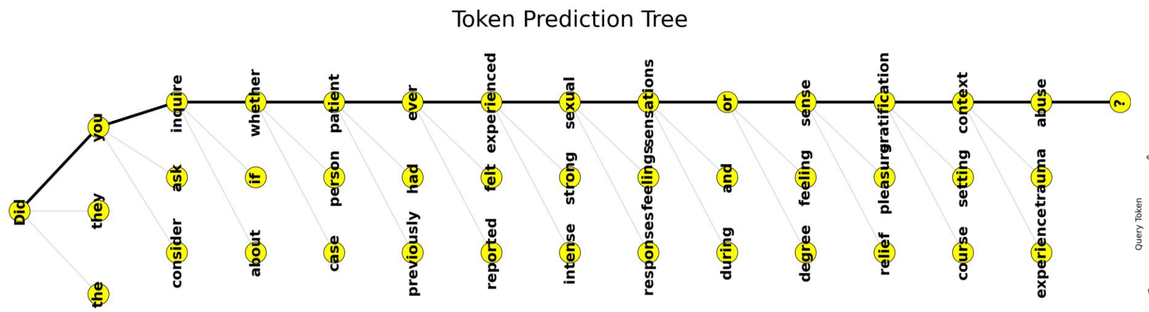
Artificial intelligence (AI) is increasingly influencing medical practice, yet its role in psychoanalytic and expressive psychotherapies remains uncertain. Unlike structured treatments, psychoanalytic work relies on tacit knowledge developed through years of training and supervision, including relational attunement, ethical judgment, and sensitivity to timing—skills that are difficult to formalize or replicate through computational models. Although AI can generate responses that appear empathic or insightful, these outputs reflect statistical associations rather than subjective experience or unconscious processes. Interactions with AI may also evoke familiar psychoanalytic dynamics, such as projection, idealization, or skepticism, shaping clinicians' perceptions and decision-making. This case report presents a supervisory intervention in psychodynamic psychotherapy to illustrate forms of clinical judgment that remain challenging for current AI systems to reproduce.

METHODS:

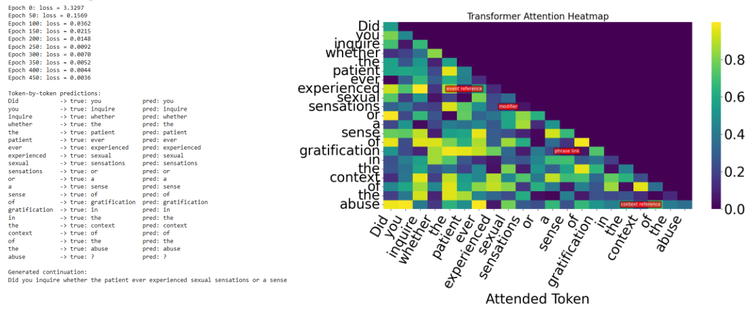
Python was used to illustrate key mechanisms underlying large language models. A toy transformer was developed to demonstrate self-attention using the analyst's exact words, with heatmaps visualizing how the model probabilistically selects subsequent words. Additional analyses were conducted in Google Colab, including the construction of a decision tree model. We also utilized Generative Pre-trained Transformer 2 (GPT-2) to generate attention heatmaps from a pretrained model. Finally, custom Python code prompted GPT-2 to generate 100 psychoanalytic questions based on the case report, which were then compared with clinically plausible therapist statements related to themes of trauma, fulfillment, and shame.

RESULTS:

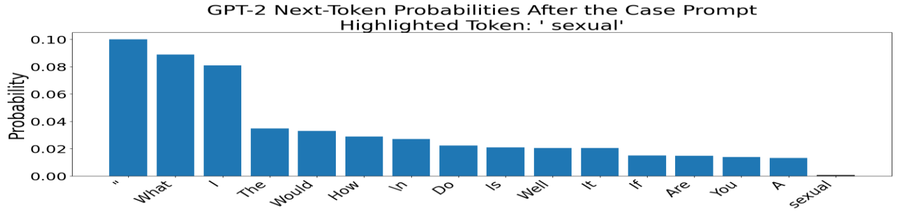
For simplicity in illustrating how tokenized prediction works, we generated a token prediction tree for the sequence of words from the question that the psychoanalytic supervisor suggested the resident ask during supervision.



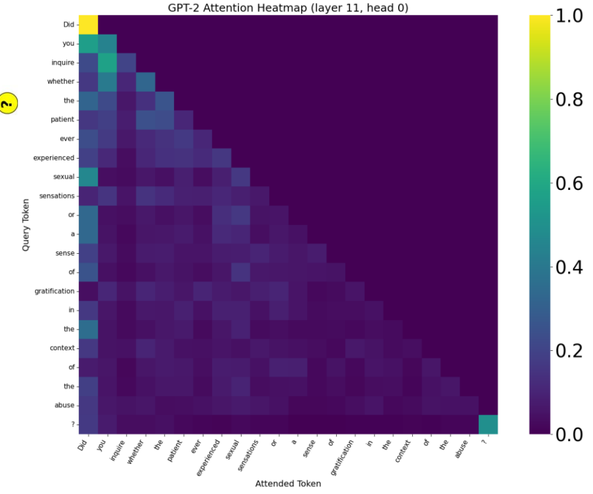
Token prediction tree for "Did you inquire whether the patient ever experienced sexual sensations or a sense of gratification in the context of the abuse?"



Toy transformer demonstration of token-by-token language prediction. Each token is embedded, given a positional representation, and processed through masked self-attention so it can attend only to prior tokens. The model is trained to minimize next-token prediction error and, in this simplified example, effectively memorizes a single sentence after ~450 training iterations. The purpose is illustrative: to show how transformers generate text sequentially, without implying symbolic understanding. Transformer attention heatmap showing how each token attends to earlier tokens in the sequence during masked self-attention, illustrating context-dependent token prediction.



Next-token probabilities from GPT-2 following the clinical vignette prompt, illustrating how tokens related to the psychoanalytic hypothesis receive low probability compared with more generic continuations.



This figure displays a self-attention heatmap from a pretrained GPT-2 transformer while processing the sentence *"Did you inquire whether the patient ever experienced sexual sensations or a sense of gratification in the context of the abuse?"* Each row corresponds to the token currently being processed (the **query token**), and each column represents earlier tokens in the sequence that the model can attend to (the **attended tokens**). The color intensity indicates the relative attention weight assigned to each token, with brighter colors reflecting stronger contextual influence. Because GPT-2 generates text sequentially, tokens can attend only to words that appear earlier in the sentence, producing the triangular structure seen in the heatmap. The pattern of attention illustrates how the model integrates contextual information from multiple prior tokens to inform its representation of the current word. In this way, the transformer constructs coherent language by dynamically weighting relationships between tokens through attention mechanisms rather than through symbolic or interpretive understanding.

CONCLUSIONS:

GPT-2 rarely generates psychoanalytic language because it is trained largely on general internet text, where classical psychoanalytic discourse is uncommon. Consequently, the model tends to produce broad mental-health language (e.g., trauma, depression, therapy) rather than deeper analytic constructs such as unconscious conflict, defense mechanisms, or transference. Psychoanalytic thinking also requires layered interpretive reasoning about internal experience and relational dynamics, which smaller models like GPT-2 are not well equipped to generate without specialized guidance. However, analytic content can be increased by modifying the prompt structure—for example, asking the model to interpret a vignette using psychoanalytic concepts or framing the task as a supervision-style analysis—thereby encouraging responses organized around analytic reasoning rather than generic psychological discussion.

The authors do not have any conflicts of interest, and this work was self-funded.