Understanding Artificial Intelligence in Psychodynamic Psychotherapy: A Clinician-Oriented Perspective

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

1.Edward Via College of Osteopathic Medicine – Virginia Campus; 2265 Kraft Drive;Blacksburg, VA 24060; 2.Bard College, 30 Campus Rd, Annandale-on-Hudson, NY 12504 ; 3.Intuitive Insight, Inc, 1765 Greensboro Station Pl, Suite 900, McLean, VA 22102; 4.Mary Washington Healthcare, 1200 Sam Perry Boulevard, Fredericksburg, VA 22401; 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:

Background: Psychodynamic psychotherapy relies on meaning-making through subjectivity, unconscious processes, symbolic interpretation, and the therapeutic relationship. As artificial intelligence (AI) tools enter mental health settings for documentation and language-based support, they raise conceptual and ethical challenges. AI systems can produce text resembling clinical interpretation despite lacking consciousness, symbolic understanding, or relational awareness, creating risk of misattributing clinical judgment to non-sentient systems. **Objective**: To provide a clinician-oriented framework explaining how AI generates clinically plausible outputs and how this differs from psychodynamic meaning-making. **Methods**: Conceptual illustrations using simplified computational models (linear regression, principal component analysis, neural networks, and language models) with synthetic datasets and psychodynamic analogies. No patient data were used. **Results**: AI models organized complex inputs into clinically recognizable patterns through statistical optimization and error minimization rather than understanding or intention. **Conclusions**: AI can mimic clinical reasoning but remains fundamentally non-interpretive. In psychodynamic practice, the primary risk is projection of meaning onto probabilistic outputs. AI may assist organization but cannot replace clinical judgment or ethical responsibility.

INTRODUCTION

Artificial intelligence (AI) is increasingly used in mental health care for documentation, monitoring, risk prediction, and digital therapeutics. Large language models can generate text that resembles clinical reasoning, psychotherapy summaries, and empathic responses. While these developments have generated excitement, they have also raised concern among psychodynamic clinicians about how AI systems function and whether their outputs represent genuine understanding. Psychodynamic psychotherapy relies on meaning-making through subjectivity, symbolic interpretation, and the therapeutic relationship—processes that cannot be reduced to computational rules. In contrast, contemporary AI systems produce outputs through statistical pattern recognition and probabilistic inference rather than intention or symbolic understanding. Without a clear conceptual framework, clinicians may oscillate between **algorithmophilia** (over-attributing insight to AI outputs) and **algorithmophobia** (rejecting potentially useful tools). This work proposes a clinician-oriented framework to clarify these differences and support informed, ethically grounded engagement with AI.

METHODS:

This study used a **conceptual, illustrative approach** to help psychodynamically trained clinicians understand how contemporary AI systems generate outputs. Rather than evaluating a specific tool, simplified computational examples were used to demonstrate core machine-learning processes and contrast them with psychodynamic meaning-making. Illustrative models included **principal component analysis (PCA), linear regression, simple neural networks, and conceptual examples of language model text generation**. Synthetic datasets were created solely for didactic purposes; **no real patient data were used**. Each computational method was paired with **psychodynamic clinical analogies** (e.g., pattern recognition, interpretive revision, thematic identification) to clarify superficial similarities and fundamental differences between statistical optimization and psychodynamic processes such as interpretation, affective attunement, and transference. The goal was to improve **conceptual interpretability and ethical awareness** when clinicians encounter AI-generated material in mental health contexts.

ILLUSTRATIVE ANALYSES:

Figure 1. PCA biplot of psychoanalytic interventions and patient responses. Six emotional response variables are reduced to two orthogonal dimensions: PC1 (Constructive vs. Defensive Outcome; 58.6% variance) and PC2 (Emotional Arousal; 38.9% variance). The plot illustrates distinct clinical profiles across interventions, with, for example, Clarification associated with high emotional arousal but an intermediate position on the constructive–defensive continuum.

Intervention	Increased Dysphoria	Improved Mood	Introspection	Externalization	Narcissistic Injury	Regression
Interpretation	25	40	55	30	20	3
Clarification	35	33	36	30	40	30
Confrontation	30	10	9	40	40	60
Praise	10	17	0	0	0	7

Table 1: Synthetic frequency data illustrating the relationship between four psychoanalytic interventions—Interpretation, Clarification, Confrontation, and Praise—and six patient responses

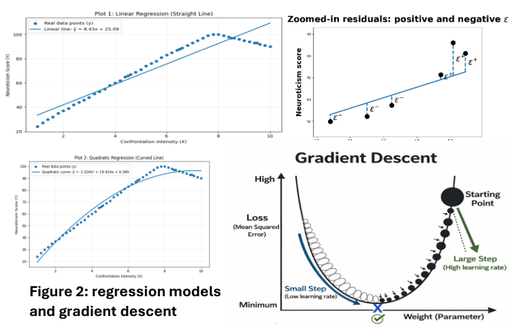
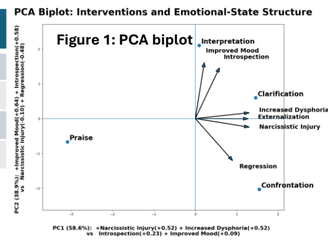


Figure 2: regression models and gradient descent

Linear vs. nonlinear modeling of confrontation intensity and neuroticism: Linear regression shows a positive association; residuals indicate prediction error; quadratic regression shows nonlinear saturation. **Gradient descent illustration:** model iteratively adjusts parameters to minimize error, with small steps converging and large steps overshooting (left). The same curve produced by Neural network (right)



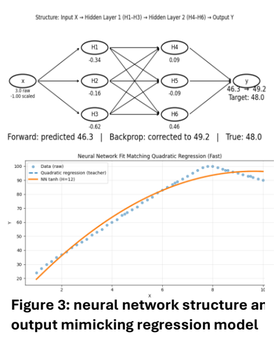


Figure 3: neural network structure ar output mimicking regression model

IMAGE RECOGNITION USING NEURAL NETWORK:

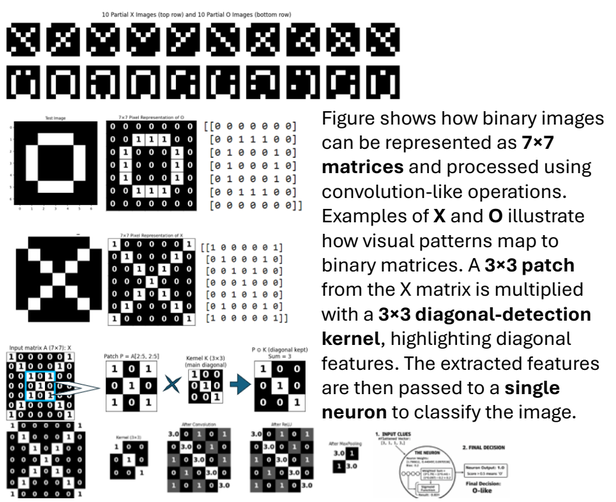
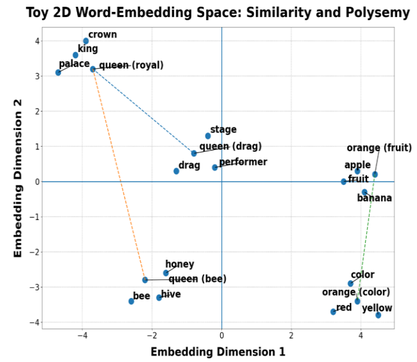
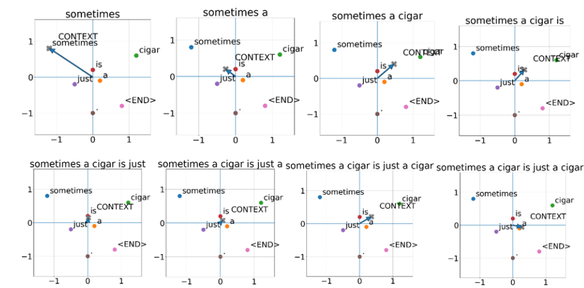


Figure shows how binary images can be represented as **7×7 matrices** and processed using convolution-like operations. Examples of **X** and **O** illustrate how visual patterns map to binary matrices. A **3×3 patch** from the X matrix is multiplied with a **3×3 diagonal-detection kernel**, highlighting diagonal features. The extracted features are then passed to a **single neuron** to classify the image.

SPEECH RECOGNITION AND PREDICTION:



In a **toy 2-D transformer**, words are embedded as points in a two-dimensional vector space.



As a sentence is generated in an **autoregressive model**, each word contributes to a trajectory through this space. The resulting vector representation of the context then points toward the **next most probable word**, guiding the model's prediction of the next token predicting Freud's famous quote "Sometimes a cigar is just a cigar."

Discussion:

- **AI will increasingly influence psychiatry and medicine.**
- **Key AI concepts can be explained in accessible terms to psychotherapists.**
- **Psychotherapy expertise is largely transmitted through tacit knowledge and supervision.**
- **AI may help study mechanisms of psychotherapy and preserve psychoanalytic knowledge if experienced clinicians contribute.**

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