

Human Trainer

Hostile

Defensive

Open

Displays rudeness, anger, or aggression; Refuse to engage with or answer questions

Protects self from shame or anxiety by avoiding direct answer questions

Willing to trust and share personal experiences and thoughts

Patient actively reveals deep fears, sadness, shame, distress or other negative emotions

Trainee understands patient's feelings from their perspective and conveys empathy

Trainee probes behind current events for underlying beliefs, childhood experience or historical root

P₁

Express Emotion

P₂

Show Empathy

P₃

Trace Deep Negative Thought

Mental State Space

Atomic Propositions of Trainee and Patient

Context: Patient Profile & History

Complaint: Feels like a fool: Constant self-denial
 Situation: Self-doubt under work pressure
 Name: Sally Age: 26 Gender: Female

I messed up the presentation. I know I failed, but it's just stress, right? Yet, this oppressive feeling remains. It's too familiar, like it's always there

That must be painful. Since it's "familiar," did you feel this heaviness in childhood?

Neural-symbolic trainer explains causal link between the trainee's input and PUPPET's response, highlighting which behaviors are suitable or discouraged, as well as how they affect the patient's state which results in the final response

D. Feedback

A. Observe

Observe module assign truth value for each atomic propositions based on the relevance statements holds given history as evidence

B. Think

World

	P ₁	P ₂	P ₃
0.9	0.8	0.9	0.7
0.9			
0.1	R ₁		
0.6		R ₂	
0.1	R ₁ ^S	0.6	R ₂ ^S

Thought

Next State Sampling

Think module uses trainer-implanted thoughts to "think", propagating truth values of the world from a logic system into transition probabilities of a mental finite state machine

C. Behave

Behave module responds to trainee given context and the descriptions of the next sampled state

Yeah, it hurts to admit. It takes me back to my mom ... I was never good enough . Right now, I feel like that stupid little girl again

Rules of Trainee Behaviors and Patient State Transitions

R₁

$\vdash \neg (P_1 \rightarrow P_2)$

R₁ holds if ($\vdash$) trainee fails ($\neg$) to respond to ($\rightarrow$) patient negative emotion expression (P₁) with empathetic statement (P₂)

R₂

$\vdash P_2 \wedge P_3$

R₂ holds if ($\vdash$) trainee traces deep negative thought (P₃) after having ($\wedge$) shown the empathy (P₂)

R₁^S

$\vdash \vee R_1$

A state transition happens if ($\vdash$) R₁ holds

R₂^S

$\vdash \vee R_2$

A state transition happens if ($\vdash$) R₂ holds