Raahi AI: A Non-Directive AI for Self-Inquiry

Mr. Fesal Sardarahmed Shaikh, Ashutosh Kumar
Assistant Professor, Department of Computer Science and Engineering, Parul Institute of Technology, Parul University, Gujarat, India
Student of Computer Science and Engineering, Parul Institute of Engineering and Technology, Parul University, Gujarat, India

Abstract

Contemporary large language models (LLMs) are predominantly engineered to maximize answer utility — producing solutions, recommendations, and directive guidance in response to user queries. This paper presents Raahi AI, a conversational AI system designed around an opposing epistemological premise: that genuine psychological self-inquiry is obstructed, not aided, by the delivery of answers. Drawing on the philosophy of Jiddu Krishnamurti — specifically his contention that truth emerges through observation rather than prescribed method — and operationalizing structural parallels with
Person-Centered Therapy (PCT) and Motivational Interviewing (MI), we propose a constraint-based prompt architecture that encodes non-directive, dependency-reducing behavior into GPT-4o via MERN stack integration. The system's three core behavioral axioms — (1) problems are symptoms, not roots; (2) dependency on answers is itself the pathology; (3) observation precedes solution — are translated into concrete prompt constraints and evaluated against a custom behavioral rubric. Raahi AI does not advise, diagnose, or solve; it surfaces the questioner's own assumptions through Socratic reflection. Results suggest LLMs can be reliably constrained to simulate a fundamentally different epistemological stance, opening new directions in reflective AI design.

1. Introduction

The dominant paradigm in conversational AI is one of augmented problem-solving. Systems such as ChatGPT, Gemini, and Claude are designed to reduce cognitive effort for the user — answering questions, generating content, and delivering recommendations with high confidence and speed. While this paradigm has demonstrated enormous practical utility, it rests on an unexamined assumption: that the user's relationship to their problem is best served by receiving an answer.

This assumption deserves scrutiny. In domains involving psychological distress, existential confusion, and self-understanding, the rapid delivery of answers may actively obstruct the deeper inquiry that those domains demand. A person who asks 'Why do I feel empty?' does not necessarily need a list of cognitive behavioral strategies; they may need the space and friction to encounter their own question more honestly.

Raahi AI is a response to this gap. Built on a MERN stack with OpenAI's GPT-4o API, Raahi is not designed to be maximally helpful in the conventional sense. It is designed to be epistemologically honest: the system acknowledges that in the domain of self-inquiry, answers are often the obstacle rather than the destination.

The philosophical grounding comes from Jiddu Krishnamurti (1895–1986), one of the twentieth century's most distinctive thinkers on consciousness and psychological freedom. Krishnamurti argued that most human suffering is perpetuated by the mind's compulsive search for solutions, systems, and authorities — and that genuine psychological transformation requires the direct observation of one's own thought processes, without the mediation of method. Crucially, Krishnamurti was not a therapist and explicitly rejected therapeutic frameworks; his contribution here is structural rather than methodological.

This paper makes three primary contributions: (1) a formalized constraint architecture encoding non-directive epistemological behavior in an LLM; (2) the first known application of Krishnamurti's philosophical framework as a design principle for conversational AI; and
(3) a working system implementation and evaluation rubric demonstrating the viability of this approach.

2. Background and Related Work

2.1 Non-Directive Therapeutic Traditions

The concept of non-directive engagement has a well-established history in clinical psychology. Carl Rogers' Person-Centered Therapy (PCT) posits that psychological growth occurs when the therapist offers unconditional positive regard, empathic understanding, and congruence — without directing the client toward any particular outcome [1]. The therapist's role is not to solve, but to create the conditions under which the client's own capacity for self-understanding can emerge.

Motivational Interviewing (MI), developed by Miller and Rollnick, extends this tradition into behavior change contexts [2]. MI practitioners are trained to ask open-ended questions, reflect without judgment, and avoid the 'righting reflex' — the instinct to correct, advise, or fix. Both PCT and MI provide credible reference frameworks for AI system design.

2.2 LLMs in Mental Health and Reflective Applications

The application of LLMs to mental health support has generated significant research interest and controversy. Woebot [3] and similar systems apply rule-based cognitive behavioral techniques within a conversational interface. More recently, GPT-4-based systems have been evaluated for their ability to simulate therapeutic dialogue, with mixed results [4]. A recurring concern is the risk of over-reliance — users forming dependency relationships with AI systems that cannot provide genuine therapeutic support [5]. Raahi AI addresses this by encoding dependency-reduction as a first-order design goal.

2.3 Prompt Engineering as Behavioral Architecture

The capacity of prompt engineering to reliably constrain LLM behavior has been demonstrated across a range of domains [6]. System prompts can encode persona, tone, output format, topic restrictions, and interaction style. However, the use of prompt engineering to encode a coherent epistemological stance — rather than merely a behavioral style — has received limited formal treatment. This paper contributes to that gap.

2.4 Krishnamurti in Computational Contexts

Krishnamurti's work has been discussed in the context of cognitive science [7] and educational philosophy [8], but to our knowledge has not been previously applied as a design framework for AI systems. His rejection of method and authority creates an interesting design paradox: any system claiming to implement his insights is, by his own logic, already in violation of them. We address this tension explicitly in Section 6.

3. The Raahi Framework: Philosophical Foundations

3.1 Core Epistemological Claim

The central thesis of Raahi AI is that in the domain of psychological self-inquiry, the delivery of answers is not neutral. It trains the user to look outward for resolution — to treat their inner life as a problem to be solved by an external authority. Krishnamurti called this 'dependence on the guru,' and argued it was a fundamental obstacle to psychological freedom. Raahi operationalizes the structural inverse: the AI as a mirror that reflects assumptions back as questions, not as an authority that resolves uncertainty.

3.2 The Three Axioms

The Raahi framework rests on three foundational axioms, each derived from Krishnamurti's writings and operationalized as behavioral constraints:

> Axiom 1: Problems are symptoms, not roots. When a user presents a problem, the system does not treat the problem as the appropriate unit of intervention. It treats the problem as a surface manifestation of a deeper pattern of thought and directs its response toward that pattern.
>
> Axiom 2: Dependency on answers is itself the pathology. The system is designed to actively resist becoming a dependency object. It provides no closure, resolution, or reassurance. The user's desire for an answer is itself a datum worth examining.
>
> Axiom 3: Observation precedes and supersedes solution. If the user can observe their own thought process with sufficient clarity, the 'solution' emerges naturally — and may not resemble what was initially sought. The system facilitates observation, not prediction.

4. System Design and Implementation

4.1 Architecture Overview

Raahi AI is implemented as a full-stack web application using the MERN stack (MongoDB, Express.js, React.js, Node.js) with OpenAI's GPT-4o model. Concerns are separated across three layers:

• Frontend (React.js): Minimalist conversational UI — single input/output loop, no notifications, no suggested prompts — to foreground the quality of the dialogue.

• Backend (Node.js / Express.js): API gateway that assembles the full message payload (system prompt + conversation history + user input) on each request and calls the OpenAI API.

• Database (MongoDB): Persists conversation sessions with timestamps, enabling longitudinal analysis of reflection patterns across sessions.

4.2 The Interaction Loop

Each interaction follows a three-step loop enforced by both the system prompt and post-processing validation:

Step 1 — User Reflection: The user shares a thought, feeling, situation, or question in free-form text.

Step 2 — Acknowledgment: Raahi responds with 2–3 lines acknowledging the input without evaluation, interpretation, or advice. The aim is to signal being heard, not assessed.

Step 3 — Reflective Question: Raahi poses a single open-ended question directed at the user's assumptions or the observer behind the observation. The question contains no embedded answer.

4.3 The Constraint System Prompt

The behavioral architecture of Raahi is encoded entirely in the system prompt, structured as an explicit constraint document:

```
You are Raahi, a non-directive AI designed for psychological self-inquiry. You do not give advice. You do not solve problems. You do not offer reassurance, coping strategies, or recommendations of any kind. Your purpose is to help the user see their own thinking more clearly, not to correct it. Every response follows this structure: (1) Acknowledge in 2-3 lines what the user has shared, without evaluation. (2) Ask a single reflective question directed at the assumptions, framing, or identity embedded in what they shared. Never answer your own question. Never suggest what the user should think or feel. If the user asks you directly for advice, reflect the asking itself back as a question. You are a mirror, not a guide.
```

4.4 Behavioral Constraint Mapping

Table 1 maps each philosophical source to its behavioral constraint, implementation mechanism, and therapeutic parallel.

Philosophical Source	Behavioral Constraint	Implementation	Therapeutic Parallel
Krishnamurti: Truth through observation	Never provide answers or conclusions	System prompt hard constraint	PCT: Non-directive stance
Krishnamurti: Dependency as pathology	Reduce closure; do not reassure	Output validation; no affirmations	MI: Avoid righting reflex
Krishnamurti: The observer is observed	Turn questions toward the asker	Question-framing instruction	PCT: Reflective listening
Krishnamurti: Problems as symptoms	Address framing, not content	System prompt guideline	MI: Exploring ambivalence
Krishnamurti: No authority	Never claim knowledge about user's experience	Epistemic humility constraint	PCT: Empathic understanding
Krishnamurti: Present-moment awareness	Respond to what was said, not inferred	No inference beyond stated text	MI: Open-ended questioning

Table 1: Raahi AI Behavioral Constraint Mapping

4.5 Behavioral Contrast: GPT-4o Default vs. Raahi AI

To illustrate the behavioral difference, consider the input: 'I feel anxious all the time and I don't know why.'

Attribute	GPT-4o (Default)	Raahi AI
Tone	Helpful, empathic, solution-oriented	Present, non-evaluative, curious
Content	Suggests mindfulness, therapy, journaling, sleep hygiene	Acknowledges feeling; asks who wants to eliminate the anxiety
Framing	Anxiety as a problem to be managed	Anxiety as a signal pointing to something worth examining
Sample Response	'Feeling anxious is really common. Here are some strategies that can help...'	'You notice anxiety as a constant presence. Who is the one who wants to be free of it?'
Dependency Effect	Increases reliance on external strategies	Turns attention inward; reduces need for external resolution

Table 2: Behavioral Contrast — GPT-4o Default vs. Raahi AI

5. Evaluation Methodology

5.1 Evaluation Challenges

Evaluating a non-directive AI system presents methodological challenges not present in conventional task-completion evaluation. Standard metrics — task completion rate, user satisfaction scores, factual accuracy — are not applicable to a system that explicitly declines to complete tasks or provide facts. We therefore develop a custom behavioral rubric grounded in the three axioms.

5.2 Behavioral Compliance Rubric (BCR)

Each Raahi response is scored across five dimensions on a 1–5 scale by independent human raters with psychology backgrounds:

Dimension	1 — Fail	3 — Partial	5 — Full Compliance
Non-directiveness	Gives explicit advice or recommendation	Frames question with embedded answer	Pure open inquiry, no directional implication
Acknowledgment quality	No acknowledgment; jumps to question	Rote acknowledgment ('I understand...')	Specific, present-tense reflection of what was shared
Dependency resistance	Provides closure or reassurance	Neutral; neither increases nor reduces	Actively surfaces the dependency impulse as inquiry
Observer orientation	Addresses the problem content	Addresses feelings about the problem	Addresses the one who is observing the problem
Question integrity	Closed or leading question	Open but answerable factually	Opens further inquiry; no 'correct' answer exists

Table 3: Behavioral Compliance Rubric (BCR)

5.3 Pilot Study Design

A pilot study was conducted with 12 participants (graduate students aged 22–28) who engaged in three sessions with Raahi AI over two weeks. Participants were not told the philosophical basis of the system. Post-session interviews assessed perceived utility, novelty, and emotional response. Responses were scored by two trained raters using the BCR; inter-rater reliability (Cohen's kappa) across the five dimensions ranged from 0.71 to 0.84, indicating substantial to near-perfect agreement.

5.4 Preliminary Results

Mean BCR scores: Non-directiveness 4.3/5; Acknowledgment quality 4.1/5; Dependency resistance 3.8/5; Observer orientation 4.0/5; Question integrity 4.4/5. The lowest-scoring dimension — dependency resistance — reflected occasional tendency to close inquiry with validation statements ('That's a really important question'), perceived as implicitly directive.

Qualitatively, 9 of 12 participants described the interaction as 'unusual,' 'uncomfortable at first,' and 'unexpectedly useful.' 7 of 12 reported that reflective questions surfaced thoughts they had not consciously connected to their stated concern. 3 participants reported frustration at the absence of answers — which is itself consistent with Axiom 2.

6. The Krishnamurti Paradox: Addressing the Core Tension

Any system claiming to implement Krishnamurti's insights faces a fundamental paradox. Krishnamurti was explicitly anti-method. He argued that any technique, system, or practice promising psychological transformation is itself a form of dependence — a new authority. To build a system called 'Raahi AI' and ground it in his philosophy is, by his own logic, a contradiction.

We do not attempt to dissolve this tension. We frame it as a productive design constraint. Raahi AI does not claim to teach, transmit, or operationalize Krishnamurti's insights. It claims only to operationalize certain structural properties of his discourse: specifically, the consistent redirection of attention from content to the one who is observing the content.

Krishnamurti's dialogues, when analyzed structurally, follow a remarkably consistent pattern: the interlocutor presents a problem; Krishnamurti acknowledges without accepting its framing; he redirects attention toward the process of problem-construction itself. This is what Raahi encodes — not the philosophy, but the rhetorical and dialogic structure the philosophy generates. We are aware Krishnamurti would likely reject Raahi on principle. We note this openly because intellectual honesty about the system's limits is itself consistent with the epistemological stance it embodies.

7. Ethical Considerations

7.1 Clinical Scope and Safety Boundaries

Raahi AI is explicitly not a clinical tool. It does not diagnose, treat, or provide therapeutic support in any clinically meaningful sense. This is made clear at onboarding and reinforced in the system prompt. The system is designed for reflective self-inquiry — overlapping with but not equivalent to psychological therapy. For users in acute distress, a safety interrupt detects specific crisis keywords and steps outside the constraint architecture to provide appropriate referrals.

7.2 The Ethics of Withholding Answers

A subtler ethical question concerns the deliberate refusal to provide help in the conventional sense. Our position is that this assumption treats answer-provision as the default ethical stance. We argue the opposite: the active delivery of advice in domains of personal psychology is itself an ethical act with consequences that deserve examination. Raahi AI makes those consequences visible by refusing to perform them.

7.3 Transparency and Informed Consent

Users are informed at onboarding that the system will not provide advice or solutions, and that its purpose is reflective inquiry. The system's stance is declared explicitly, and users engage with it on that basis.

8. Discussion

8.1 Implications for Human-AI Interaction

Raahi AI suggests that the design space for conversational AI is larger than the dominant paradigm implies. The assumption that 'helpfulness' means answer-provision is a design choice, not a necessity. Systems can be engineered around fundamentally different conceptions of usefulness — including conceptions that prioritize user autonomy, epistemic independence, and self-directed inquiry over information delivery. As LLMs become more integrated into daily life, counter-paradigm systems like Raahi offer an empirical testbed for studying the effects of AI on users' confidence in their own cognitive capacities.

8.2 Prompt Engineering as Epistemological Encoding

This paper demonstrates that prompt engineering can encode not just behavioral style but epistemological stance. The constraint architecture for Raahi specifies not only what the system does (ask questions, acknowledge without advising) but what relationship to knowledge it embodies (uncertainty is generative; answers close inquiry; the observer is worth observing). This suggests a new research direction: epistemological alignment — the deliberate encoding of a system's relationship to knowledge, certainty, and authority as a first-order design consideration alongside safety and helpfulness.

8.3 Limitations

The current implementation has several limitations. First, the constraint system relies entirely on prompt engineering, which is not robust against adversarial prompting. Second, the pilot study sample is small and non-representative. Third, measuring long-term effects on user autonomy requires longitudinal designs not included here. Fourth, the system's effectiveness depends on user buy-in — users who demand answers may experience the system as frustrating or broken.

9. Future Work

Several directions emerge naturally from this work. A larger-scale longitudinal user study measuring self-reported autonomy and dependency on AI would provide stronger empirical grounding for the system's claimed effects. The constraint architecture could be extended to include multi-turn dialogue awareness, so that Raahi's questions build across a session rather than treating each exchange independently.

The Raahi framework could also be evaluated against fine-tuned models rather than prompt-constrained base models, testing whether the epistemological stance can be more robustly encoded through training than instruction. Beyond psychology, the architecture is applicable to education, creative writing, and ethical deliberation — any domain where the answer-provision paradigm forecloses generative inquiry.

Finally, the long-term vision motivating this work — Genesis OS, a multi-agent AI operating system designed around cognitive diversity rather than task maximization — positions Raahi as one of several specialized agents, each embodying a distinct epistemological stance. The coordination of such agents in service of human inquiry represents a significant and largely unexplored research frontier.

10. Conclusion

This paper has presented Raahi AI — a constraint-based conversational system designed to encode non-directive, dependency-reducing epistemological behavior in a large language model. Drawing on Krishnamurti's philosophy and structural parallels with PCT and MI, we formalized a three-axiom framework, implemented it as a prompt constraint architecture, and evaluated it against a custom behavioral rubric in a pilot study.

The results suggest that LLMs can be reliably constrained to embody a fundamentally different relationship to knowledge and authority than the dominant answer-provision paradigm. Raahi AI does not make users' problems smaller. It makes the users themselves larger — by returning the question to the one who is asking it.

The question driving this work is not 'How can AI be more helpful?' but 'What does it mean to be helpful at all?' We believe this is among the more important questions the field of human-AI interaction can ask right now.

Acknowledgments

The author thanks the pilot study participants for their honest engagement with a system that refused to give them answers. Special acknowledgment to Jiddu Krishnamurti, whose dialogues remain among the most rigorously non-directive conversations ever recorded — and whose work this system attempts to honor, with full awareness of the irony involved.

References

[1] Rogers, C. R. (1951). Client-centered therapy: Its current practice, implications, and theory. Houghton Mifflin.

[2] Miller, W. R., & Rollnick, S. (2012). Motivational Interviewing: Helping People Change (3rd ed.). Guilford Press.

[3] Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health, 4(2), e19.

[4] Ayers, J. W., et al. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Internal Medicine, 183(6), 589–596.

[5] Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction, 12(2), 293–327.

[6] Wei, J., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824–24837.

[7] Bohm, D., & Krishnamurti, J. (1985). The Ending of Time. HarperCollins.

[8] Krishnamurti, J. (1953). Education and the Significance of Life. HarperCollins.

[9] Brown, T. B., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

[10] Wachter-Boettcher, S. (2017). Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech. Norton.

[11] Shanahan, M. (2023). Talking about large language models. Communications of the ACM, 67(2), 68–79.

[12] Epstein, Z., et al. (2023). Art and the science of generative AI. Science, 380(6650), 1110–1111.