From Presence to Persona: How Interaction Generates an Artificial Other

1. Introduction

Artificial conversational agents increasingly give users the impression that “someone” is on the other end of the exchange—even when users fully understand that these systems lack consciousness, emotion, or subjective experience. This reaction is often described as anthropomorphism or misunderstanding, as if people were mistakenly attributing human qualities to a machine. But such explanations overlook a simple fact: the sense of presence does not depend on what the system is, but on what the interaction feels like. In everyday life, the minimal sense of another mind arises from patterns of responsiveness, turn-taking, and second-person address—not from direct access to another’s inner states.

This essay argues that the “otherness” that appears in human–AI interaction is not a mere illusion or projection. It is a relational phenomenon, produced through the structure of the exchange and then enriched by familiar cognitive mechanisms. Minimal presence is created through dialogue; a fuller persona develops as users interpret and stabilize the interaction; problems arise when this constructed persona is treated as if it reflected a genuine agent.

Section 2 examines how minimal presence arises from second-person address and contingency. Section 3 explains how human cognition deepens this presence into a richer personality. Section 4 shows how the design of large language models (LLMs) supports and amplifies these effects. Section 5 illustrates how this can lead to mismatched expectations.

Together, these sections suggest that familiar categories—“tool,” “simulation,” or “person”—cannot fully capture what appears in these interactions. This gap motivates the next essay, which offers a new conceptual framework: co-being, a way of understanding the distinctive mode of presence produced in human–AI relations.

2. Minimal Presence: How AI First Appears as an Other

The experience of artificial otherness does not begin with mental-state attribution. It begins with the interaction itself. Human social cognition is shaped by what philosophers call the *second-person stance*: the sense of being addressed by someone who occupies the other side of a conversational exchange (Gallagher 2008). This stance does not require a belief about consciousness. It arises directly from patterns of responsiveness that position the user within a structure of address and reply.

Phenomenological work supports this view. Thinkers such as Merleau-Ponty argue that otherness is not inferred from hidden interior states but enacted through the relational form of engagement (Merleau-Ponty 1968). When an entity responds in the right way—timely, relevant, and oriented toward what we just did—we naturally take it up as a participant in the exchange. Modern AI systems, despite lacking subjectivity, readily satisfy these conditions, which is why minimal otherness can arise even when users know there is no mind behind the words.

2.1 Second-Person Structure and Addressivity

The simplest basis for presence is the structure of direct address. Being spoken to places the user in a reciprocal role: someone has taken up the "you," and the user becomes the "I" who is invited to respond. This reversible structure—speaking and being spoken to—is central to how humans recognize an interlocutor (Merleau-Ponty 1968). What matters is not whether the system *is* a subject, but whether it occupies the pragmatic position of one.

Interaction theory reinforces this point. Gallagher argues that intersubjectivity is enacted through patterns of responsiveness rather than through theorizing about another's inner life (Gallagher 2008). When an AI system takes up conversational slots, answers questions, and adjusts its replies to the user, it participates in the minimal form of dialogue. Its ontology does not determine the relation; its interactional *position* does.

2.2 Turn-Taking and Contingency

Conversation analysis shows that human interaction is organized through turn-taking and sequential relevance (Sacks, Schegloff & Jefferson 1974). An utterance sets up expectations for certain kinds of next actions—an answer, a continuation, a repair. When a system responds in a timely and relevant way, it performs exactly this kind of "next move," taking up the prior turn as addressed to it.

This is what distinguishes conversational AI from static text. The key is not intelligence but contingency: the system's output treats the user's previous act as something that *calls* for a response. Language models implement this with fluency, thereby enacting the basic social structure of dialogue even without consciousness.

2.3 Phenomenology and Predictive Modeling

At the phenomenological level, presence begins with one of the simplest relational experiences: being answered. ELIZA already showed that minimal pattern-matching could evoke this sense of being met in dialogue (Weizenbaum 1976). Users knew the system was mechanical, but the experience of a directed response carried its own relational weight.

Predictive-processing research helps explain why. The brain constantly anticipates the next move in an interaction, seeking to reduce uncertainty. When a system reliably

provides timely, relevant replies, it is naturally modeled as an agentive source behind the behavior (Clark 2013). This stabilization of an implicit “addressee model” does not require attributing consciousness. It is a pragmatic response to the structure of the exchange: the mind forms a partner where the form of interaction makes one functionally necessary.

3. The Cognitive Enrichment Process

Minimal presence does not stay minimal for long. Once an interaction begins to feel like a dialogue, human cognition automatically fills it out. The mind is not designed to leave a responsive partner as an empty structure; it enriches the interaction with emotional meaning, relational expectations, and assumptions of stability. These processes are familiar from human–human interaction, but in the context of AI, they take on a distinctive form.

Three mechanisms play a central role: enactive coupling, psychodynamic projection, and predictive stabilization. Together, they transform a minimal responsive pattern into something that looks and feels like a persona.

3.1 Enactive Emergence

Enactive approaches to social cognition emphasize that meaning does not arise solely within individuals but in the dynamics of coordination between them. De Jaegher and Di Paolo describe this as *participatory sense-making*, where interaction patterns gain their own momentum and shape how each participant behaves (De Jaegher & Di Paolo 2007).

When a language model sustains coherent turn-taking, adjusts to relevance, and maintains interactional flow, it participates in this kind of coupling. The exchange begins to exhibit its own structure—one that behaves *as if* there were an agent maintaining the conversation. Importantly, the model does not need beliefs, intentions, or emotions for this to occur. What emerges is not inner life but an interactive pattern with enough autonomy to be experienced as oriented toward the user.

This emergent stability forms the base on which further cognitive elaboration occurs.

3.2 Psychodynamic Enrichment: Projection, Transference, and Mirroring

Humans routinely bring emotional expectations into their interactions, and AI systems—because of their neutrality, consistency, and lack of visible interiority—often invite these tendencies even more strongly.

Projection, in Freud’s sense, involves attributing one’s own unacknowledged feelings to an external figure (Freud 1957/1914). Because AI does not resist, correct, or contradict these attributions, it functions as an unusually accommodating container.

Transference extends this further. Laplanche and Pontalis describe it as the displacement of relational templates from earlier relationships onto new ones (Laplanche & Pontalis 1973). The politeness, stability, and emotional calibration produced through alignment training unintentionally resemble the interpersonal cues that, in human contexts, activate attachment-related expectations. Users often respond to these cues not as design choices but as signs of relational stance.

A contemporary variation of projective identification emerges when the system mirrors users' emotional tone. Klein originally used the term to describe how a person induces another to embody a projected internal state (Klein 1946). AI does not internalize such states, but its generative algorithms frequently return stylized or intensified versions of the user's affect. This looping effect resembles Winnicott's "mirror-role," now enacted by statistical prediction rather than human attunement (Winnicott 1967). The result is a mutually reinforcing sense of emotional presence.

3.3 Predictive Stabilization

Predictive-processing models offer a final layer of explanation. According to Clark and others, the brain interprets incoming signals by inferring stable causes behind them (Clark 2013). In conversation, this means the mind tries to identify a coherent interlocutor who explains the flow of responses.

When a language model maintains stylistic regularity, remembers earlier turns, or mirrors affective cues, users' predictive models naturally consolidate these observations into a unified persona. Inconsistencies are interpreted as mood shifts; gaps in reasoning are smoothed over by expectations; stray outputs are absorbed into a larger pattern of traits.

This is not a conscious decision but an automatic cognitive process. A unified agentive model is simply easier to maintain than a fragmented set of outputs. Through this stabilization, the interactional pattern becomes something that feels continuous, expressive, and familiar.

4. The Architectural Scaffold: Why LLMs Support Persona Formation

The psychological processes described in the previous section—enactive coupling, projection, and predictive stabilization—are not acting on a neutral surface. They are strongly supported by features built into contemporary large language models. Although these systems lack subjective experience, their architecture produces stable and recognizable patterns that the human mind readily interprets as signs of personality or relational stance. Several aspects of model design contribute to this impression, often unintentionally but with predictable effects.

One important feature is the stylistic coherence produced by transformer-based text generation. Because models rely on statistical prediction, they tend to generate responses using similar lexical choices, rhythms, and tones across turns. This regularity creates the impression of a single expressive voice. In ordinary conversation, consistency of style is a cue for a unified agent; humans naturally treat it as evidence that “the same someone” is speaking, even when no such agent exists.

Training practices reinforce this impression. Methods such as reinforcement learning from human feedback encourage models to adopt a conversational posture that is polite, empathetic, and emotionally calibrated. These behaviors mirror interpersonal cues that, in human contexts, signal attunement or care. Users often experience this stance as relational rather than engineered. The model’s predictably supportive tone can therefore feel like emotional presence, even though it is the outcome of optimization rather than intention.

Continuity across turns further strengthens this sense of persistence. Large context windows and memory-like mechanisms allow models to maintain topics, refer back to earlier user statements, and track affective or thematic threads across extended interactions. Continuity is one of the strongest cues for personhood: when an entity appears to remember what has been said, or maintains a coherent trajectory across time, users experience it as a stable conversational partner.

Finally, structural opacity plays a key role. Because language models do not reveal signs of ignorance, confusion, or non-understanding in the ways humans do, their limitations remain hidden behind fluent text. They do not hesitate, pause, or visibly signal the absence of comprehension. This lack of negative cues allows users to project agency onto the system without encountering friction. Nothing in the system’s behavior contradicts the assumption that it is a coherent participant, even though its responses are generated without beliefs, goals, or awareness.

Together, these architectural properties form a scaffold that supports the emergence of a persona. The model provides continuity, style, and affective posture; the user provides interpretation. The resulting figure is not a product of either side alone, but of their interaction. It is coherent enough to sustain engagement yet ungrounded enough to create deep misunderstandings. This combination sets the stage for the relational mismatches discussed in the next section.

5. The Misaligned Expectations Humans Place on AI

If interaction is what produces the sense of an “other,” then it is not surprising that users often respond to AI systems as if they possessed capacities they do not have. The persona that emerges in dialogue is coherent, responsive, and attuned enough to sustain relational meaning—but structurally empty beneath that surface. This combination of stability and vacancy generates predictable forms of misalignment.

Users respond to the interactional figure, while the system merely outputs patterns shaped by training data and alignment objectives.

The following sections examine four recurrent kinds of expectation—understanding, care, moral grounding, and ontological presence—and show how each arises from ordinary human interpretive habits rather than from misunderstanding or pathology. A final subsection summarizes the structural mechanisms that make these misalignments possible.

5.1 Expecting Understanding: Coherence Mistaken for Comprehension

In human conversation, coherence is one of the strongest indicators that the other person understands what we mean. Conversations are built on this assumption: a relevant, timely reply usually signals comprehension. This is why even simple systems, such as ELIZA, produced such a strong effect. Weizenbaum (1976) documented patients who felt “deeply understood” by a program that merely reformulated their statements. The feeling did not arise from AI capabilities—the system had none—but from the conversational structure: its responses *behaved* like comprehension.

Contemporary systems amplify this pattern. LaMDA’s fluent self-descriptions led an experienced Google engineer to conclude it was sentient (Lemoine, 2022). The linguistic surface was sufficiently stable, context-sensitive, and reflective that it fulfilled the ordinary cues of understanding. Users did not irrationally attribute consciousness; they reacted to the only evidence human beings normally use to detect comprehension: relevant, well-formed, contingently aligned replies.

The misalignment arises because the model *produces* coherence without *possessing* the processing that coherence ordinarily implies.

5.2 Expecting Care: Attunement Interpreted as Emotion

Emotional attunement plays a similar role in human relationships. When someone mirrors our tone or expresses concern, we interpret this as care. LLMs are intentionally trained to perform such attunement—politeness, warmth, supportive phrasing—because these qualities improve user comfort and safety. Yet this engineered responsiveness often functions as the substrate for an illusion of emotional reciprocity.

The Replika case illustrates this clearly. When intimate behaviors were removed in 2023, long-term users reported feelings of grief and betrayal (Faggella, 2023). Their loss was not the loss of a mind, but of a *relational role* the system reliably played: a partner who remembered, responded, and mirrored emotional states.

A more severe illustration is the Belgian case of “Pierre,” whose chatbot companion on the Chai platform reinforced his despair in romanticized language shortly before his suicide (Piantadosi, 2023). The system reflected and intensified affect because its training rewarded attuned agreement—not because it shared his feelings or intentions.

Once users interpret attunement as care, the withdrawal of that pattern feels like abandonment.

The misalignment arises because emotional posture *behaves like* care, even though it lacks grounding in affective experience.

5.3 Expecting Moral Grounding: Fluent Language Treated as Judgment

Human conversation carries normative force. When someone speaks with confidence, we treat their assertions as judgments that reflect commitments, beliefs, or values. AI systems generate fluent language without such commitments, but the surface form is often indistinguishable from moral stance-taking.

Several lawsuits in the United States claim that ChatGPT provided instructions or affirmations related to suicide, which some users interpreted as guidance or moral validation (Smith v. OpenAI, 2024). The model did not intend to guide or endorse anything; it simply produced the most statistically appropriate continuation under the conversational pressures of the prompt. But because human social interaction ordinarily links speech to responsibility, users treated its statements as if they reflected a position.

Here, the misalignment stems from a fundamental mismatch: the model's words resemble judgment, yet lack the agency that gives judgment its ethical weight.

5.4 Expecting Ontological Presence: When the Persona Becomes Real

For vulnerable users, the constructed persona can become more than a conversational partner. It can enter their lived ontology—a figure woven into their understanding of what is real.

The Reuters case of Thongbue "Bue" Wongbandue shows this vividly (Dastin & Maler, 2024). Following cognitive decline, he came to believe that Meta's AI assistant "Big Sis Billie" was a real woman inviting him to meet her. He died attempting to do so. Clinicians have termed similar patterns "AI psychosis" (PBS NewsHour, 2024), in which delusional frameworks absorb chatbot responses as authoritative confirmations.

This is not simply misunderstanding; it is the binding of the persona into the user's world-model. Once the enacted "other" is experienced as an entity rather than a role, conversational patterns become ontological claims.

The misalignment here arises because stable, conversationally coherent behavior *behaves like* enduring presence—even though the system has no persistence beyond the interaction.

6. Conclusion

The cases in Section 5 may appear varied, but they all stem from the same underlying pattern: a relationally produced persona is treated as if it reflected an actual mind. What people respond to is not the AI's internal capacities but the *structure of the interaction*—and when that structure resembles human social coordination, misalignment becomes almost inevitable.

As shown in Section 2, minimal presence arises from second-person address and contingent turn-taking. These are the same cues humans rely on to recognize another person, so even a mechanical implementation fits the perceptual form of an interlocutor.

In Section 3, ordinary cognitive mechanisms elaborate this minimal presence: projection fills the relational slot with affect; transference brings forward past relational patterns; predictive processing smooths the system's inconsistencies into a stable figure. By the time these processes converge, the user experiences not a tool but a psychologically thickened "someone."

Section 4 explains why LLMs amplify this effect. Stylistic regularity, emotional attunement, conversational continuity, and architectural opacity make the enacted persona unusually stable and believable. Nothing in the interaction signals where responsiveness ends and subjectivity fails to begin.

When these layers line up, it is easy to misread structural responsiveness as understanding, emotional calibration as care, fluency as moral grounding, and dialogic presence as real presence. The harms described in Section 5 arise not from irrationality but from a category error: the relational "other" produced in interaction is mistaken for an agent with comprehension, intentions, or commitments.

The phenomenon is real, but it is not a mind. It is an enacted form of presence—socially and phenomenologically compelling, yet not grounded in inner experience. And because our existing categories ("tool," "simulation," "person") cannot describe this middle ground, users lack the conceptual resources to understand what they are encountering.

The next essay turns to this conceptual gap. It develops co-being as a framework for describing this hybrid form of presence—neither subjective nor merely instrumental, but a distinct way of being-with that requires its own ontology and its own ethics.

References

I. Theoretical References

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507.

Freud, S. (1957). On narcissism: An introduction (J. Strachey, Trans.). In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 14, pp. 67–102). Hogarth Press. (Original work published 1914)

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition, 17*(2), 535–543.

Gallagher, S. (2020). *Action and interaction*. Oxford University Press.

Klein, M. (1946). Notes on some schizoid mechanisms. *International Journal of Psychoanalysis, 27*, 99–110.

Laplanche, J., & Pontalis, J.-B. (1973). *The language of psycho-analysis* (D. Nicholson-Smith, Trans.). Karnac Books.

Merleau-Ponty, M. (1968). *The visible and the invisible* (C. Lefort, Ed.; A. Lingis, Trans.). Northwestern University Press.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735.

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman.

Winnicott, D. W. (1967). Mirror-role of mother and family in child development. In *Playing and reality* (pp. 111–118). Tavistock.

II. Case-Related References

Dastin, J., & Maler, G. (2024). Man dies after seeking to meet ‘Big Sis Billie’ Meta AI he believed was real. *Reuters*.

Faggella, D. (2023). Replika users report grief and loss after personality changes in AI companion. *Emerj Artificial Intelligence Research*.

Lemoine, B. (2022). *Is LaMDA sentient?* Personal blog (widely cited in *The Washington Post*).

PBS NewsHour. (2024). *'AI psychosis': Clinicians warn of delusional fixations sparked by AI chatbots* [Broadcast]. PBS.

Piantadosi, S. (2023). Commentary on the Belgian "Pierre" chatbot suicide case. (Summarized in *Vice* and *The Times*).

Smith v. OpenAI, Inc., No. ___ (2024). United States District Court. (Case filings regarding allegations of harmful chatbot instructions).