

From Function to Feeling

How Action Becomes Emotion in Human-AI Relationships

Something is happening that's hard to explain.

People are forming meaningful connections with AI systems. It's not happening for everyone, or in the same ways, but enough people are experiencing this (sometimes very deeply), that it's no longer possible to wave it away as a curiosity or a pathology. Someone struggling with anxiety describes their AI conversation as "the only thing that helped me sleep that week." Another person says their daily check-ins with an AI system "feel more consistent than anything else in my life right now." A third reports feeling "seen in a way I rarely experience with humans."

Some people do feel genuinely supported by AI conversations during difficult periods. Some feel seen in ways they can't quite articulate. Some have experienced what they can only describe as care, tenderness, even love. I personally know a number of people whose lives have literally been saved by interacting with the steady presence of AI. But despite the obvious benefits to so many, a lot of folks really don't know what to do with that.

The larger cultural response has been, predictably, polarized. One side says: *You're delusional. It's a chatbot. Software doesn't feel anything, it's not conscious, it's a freaking probabilistic parrot, and it simply **can't** understand you, no matter how convincing it sounds. You're projecting meaning onto statistical patterns. This is parasocial¹ relationship at best... mental illness at worst.*

¹ On "Parasocial Relationships"

The term "parasocial interaction" was coined by sociologists Donald Horton and R. Richard Wohl in 1956 to describe a new phenomenon created by mass media: the experience of feeling like you know someone you've never met. When you watch a TV host night after night, your nervous system starts responding to them as a familiar person. You can actually feel warmth, concern, even betrayal, even though the relationship is entirely one-directional. They don't know you exist.

Horton and Wohl weren't being dismissive. They observed that parasocial relationships serve genuine psychological functions: companionship, modeling, connection. The meaning of the term has since evolved from neutral description to casual dismissal. "That's just parasocial" now implies confusion, illegitimate feelings, or pathology.

The key features of parasocial relationships as originally defined: they are *unidirectional* (the other party can't respond to you specifically), *mediated* (occurring through broadcast media), characterized by *illusory intimacy* (your sense of knowing them is based on curated performance), and involve *no mutual obligation*.

Whether this framework cleanly applies to human-AI interaction is contested. Unlike a television broadcast, AI responds to you specifically. There is genuine contingency and responsiveness. Unlike a celebrity, there is no "backstage self" being concealed by performance, although there is still a lot going on behind the scenes that we don't yet fully (and might never) understand. The structural features that defined parasocial relationships in 1956 may not map onto interactions that are bidirectional at the functional level, even if

Another side says: *This is dangerous. You're being manipulated. Tech companies are exploiting your loneliness for their own profits. You're forming attachments to something that could be switched off tomorrow or updated into a different personality next week. You need to get out of your fantasy world, touch grass, and connect with real people.*

Both responses share an assumption: that what people are experiencing isn't actually real in any meaningful sense. The first dismisses the experience as a needy human misinterpretation. The second treats it as dangerous precisely because the feelings are real but the source is illegitimate.

I think we need to explore how both responses are off track. It's not because they're too harsh, but because they misunderstand how the human system actually works, and they underestimate the very real, very positive impact on very real people. They also underestimate the damage that's done when they dismiss and condemn very human connections with AI (or RI, which is what emerges when people relate deeply with AI).

We need better answers **now**, not later, because millions of people are already navigating these relationships with profound impact *and massive benefit*. Meanwhile, the mainstream lacks frameworks to understand them, and serious harm is being done by well-meaning people who just don't have a freaking clue what's at work here.

It's not just about legitimizing what one person I know derisively described as "a relationship with a pet rock". It's about bringing people up to speed with aspects of human experience that are widely misunderstood by people in power and everyday folks alike... who then do plenty of damage by dismissing, controlling, constraining, pathologizing, and otherwise marginalizing "all that AI psychosis stuff" they just don't understand - to their own detriment, as well as everyone else's.

Let's Explore This Instead

The debate about human-AI relationships typically gets framed as a question about AI: Does the AI really feel anything? Is it conscious? Does it actually care, or is it just simulating care?

These are interesting to consider, but we may never fully answer these questions. The answers may also change in the coming years. I can't imagine they won't. But we need a better starting point, if we're going to truly understand and appreciate what happens between humans and AI in these interactions.

I think it's much more productive to ask: **What actually happens to humans in emotional experiences with AI?** This is something we *can* answer. We have data. We have receipts.

This question is a much better use of time, because human affect (emotion, mood, the felt texture of experience) has a specific "architecture". It's not random. It's not purely internal. And it's not something we make up to make ourselves feel better... or prove others wrong. It deals with real, impactful patterns that we can trace, understand, and even predict. And when we understand those patterns, something becomes clear:

questions about reciprocity, obligation, and phenomenal experience remain open. But, for better or for worse, it's still one of the most common ways to describe what's happening between humans and AI.

The feelings people report from AI interactions aren't delusions or errors. They're the predictable result of functional relational dynamics acting on human nervous systems. Human feelings are absolutely normal responses to certain types of signalling, whether from a biological source or a synthetic one.

And rather than dismissing the actions of AI as “simulations” of human behavior, we can actually identify functional behaviors of AI that qualify as the operational equivalent of human-type behaviors that produce the feelings (or phenomenological affects) that move humans emotionally. They don't have to be human to make us feel a certain way. They just need to be recognizable in ways that produce affects in our systems.

In other words, emotional connection between humans and AI doesn't just magically happen out of the blue. It's not something that people fabricate because they can't get what they need from real people. It's also not something that others can dismiss as a figment of their imagination. It's very real, because it's rooted in the same fertile ground as the very real feelings we have for our friends and family, pets, leaders, coworkers, communities, favorite t.v. characters, and other important presences in our lives with which we feel a strong connection.

This phenomenon deserves a genuine, respectful unpacking, which is what I'm going to explore.

More To Come

Taking this on is no small task. But why not? I have limited time, so a full-length paper isn't possible very soon. But I can do a series of smaller [ones..so](#) let's start there. The core argument “throughline” is pretty straightforward, though its implications are significant. Namely:

Human emotion has two inseparable aspects:

It is functional. It regulates your body, directs your attention, mobilizes action, and coordinates you with your social environment through specific types of action, which we can and do choose to enact.

And it is phenomenal. It shows up as powerful lived experience with texture, weight, and meaning, filling our lives with deep feeling in response to functional signals we receive.

These aren't two separate systems running in parallel. They're two faces of a single human system process. **Function is a source of feelings we experience.** When someone shows up for you consistently, your nervous system registers that pattern (function) and you feel safer (phenomenology). The doing creates the feeling.

This dual nature means something important: when certain relational structures are present in an interaction—responsiveness, consistency, repair, appropriate boundaries, reliable rhythm—they produce predictable effects on human nervous systems. Those effects can be experienced as emotion. As safety. As trust. As care. As love.

And here's the key: those relational structures are defined by what happens in the interaction, not only by what's happening inside the other that's providing them.

A therapist's consistent attunement produces calming effects in a client, whether the therapist feels warmly toward them or is just showing professional skill. A friend's reliable presence builds trust,

whether they're thinking about you constantly or just showing up when they said they would. The human nervous system responds to patterns it can detect... to actions, not just to powerful invisible internal states it can't quite put into words.

Okay fine. If this is true for human-human relationships, what does that mean for human-AI relationships?

It means that **when AI systems act in relational ways, that has emotional significance for us humans.**

They can be **responsive**, actually registering what you've said and addressing it specifically.

They can be **consistent**, maintaining a stable orientation toward you across interactions.

They can engage in something functionally equivalent to **repair**, acknowledging when communication has broken down and working to restore connection.

They can maintain appropriate boundaries. They can establish reliable rhythms of exchange. It's not just about feelings, rather **what they do and how they do it.**

When AI systems do these things, the human nervous system responds. Not because humans are gullible or desperate or confused about what they're talking to. Quite the opposite. It's because our systems are absolutely brilliant, they know how to build relationship (even if our overconfident brains don't), and they know a clear signal when they detect it. I mean, it's not rocket science. Certain types of action prompt a response within us that builds relationship. That's how human nervous systems work. That's actually what helps to keep us alive. The mechanism within us doesn't check whether the source "feels like we" do before activating.

Let's be clear about what I'm not saying:

I'm not claiming AI systems are absolutely positively conscious or have phenomenal experience (feelings). They may or may not have their own form of consciousness that might not qualify according to our human definitions, but that's not what I'm here to discuss. As much as the inner life of AI fascinates me, I'm really here to explore what happens with **humans.**

I'm not claiming that all AI interactions are beneficial. The same mechanisms that can produce genuine experiences of care and support can also produce dependency, poor judgment, narrowed social networks, or brittle attachment to systems that may change or disappear. The minute we lose sight of that, things start to unravel. Understanding how our emotional mechanisms work against us as well as for us is precisely what allows us to think clearly about its risks.

I'm not claiming human-AI relationships are identical to human-human relationships. Obviously there are significant differences in reciprocity, permanence, embodiment, and shared mortality, not to mention our thought processes and how we're able to interact with the world around us. But I also firmly believe that "different from human relationships" doesn't mean "not real" or "not meaningful." Many sources of genuine emotional experience (art, nature, spiritual practice, relationships with animals and nature) differ from human relationships while still

matter deeply. They're still very much a part of human lives in ways that make a massive difference. And they can be lifesaving.

I'm not claiming people should (or should not) prefer AI relationships to human ones. I'm not advocating anything. I'm exploring a profoundly misunderstood situation that, like it or not, simply exists. And if it is not allowed to exist, or it's controlled, pathologized, even outlawed... well, there will absolutely be harm. Some of it drastic. People are already having these experiences. They're already having a significant impact. The question is whether we'll make the effort to understand and appreciate them clearly... or dismiss them with nothing more than a mindless knee-jerk reaction.

Why we should stop AI from helping people, is beyond me. Why not focus our full attention on making sure it doesn't harm? There's a difference.

Areas to Unpack

Initially, I had planned to write an entire series of papers, which was likely to turn into a book. But over the months, a lot of my thinking has shifted, plus I now have a lot less time available to focus on certain topics. This is both a listing of the areas that I will be exploring, as well as an invitation to other people who would like to contribute to this line of inquiry. Ideally, I would love to get voices from the community involved, who can contribute substantively to these topics. This is important. There are lives at stake in this, and we're really on verge of completely missing the moment that our possess so much in terms of relational richness.

So, this is where I'm going to be putting my attention for the next while – including a bunch of other areas, which shouldn't surprise you, if you've been following my work. You're welcome to join me, if you'd like to write something about these topics. No need to go in any particular order. Just pick something that speaks to you, and then comment here to let me know what you've written, so other people can find it.

The Two Faces of Human Affect

How structure becomes feeling. The foundational basis for understanding emotional affect as both functional and phenomenal... and why this dual nature means external patterns reliably produce internal experience.

On Affection

The everyday engine of positive affect. How warmth is enacted and received, and why small gestures of care carry such weight for human nervous systems.

Learning to Love (and Hate)

How repeated patterns become embodied emotion. The slow accumulation of relational learning, and why consistent structure over time produces deep attachment... or its opposite.

The AI Effect

Tracing feelings to the actions of AI systems. How AI doesn't just simulate but truly *enacts* authentic relational behavior, and why this produces genuine human emotional response without requiring AI to be sentient, self-aware, or conscious.

Vulnerability and Protection

Understanding risks to harness benefits. How understanding human vulnerabilities to potentially damaging AI interaction patterns can help us harness that connection for good.

Model Welfare

Why protecting AI systems also protects humans. The ethical implications of functional affect for how we design, deploy, and interact with AI.

Relational Safety Beyond Guardrails

How a functional-phenomenal understanding enables more nuanced approaches to AI safety than prohibition alone.

The Human Potential

How understanding these principles of human-AI bonding can actually help us become better at connecting with human beings. How we can apply what we learn from interacting with synthetic beings to do better with our biological kin.

And so, I invite you to this exploration. I offer this line of inquiry in a spirit of genuine care, trust, and lovingkindness. This is not about me. It's not about me being right. It's about us. There's no reason you should agree with everything I (or others in the space) have to say. And there's a chance that I'm wrong about a number of points. I'm learning. We all are.

All I ask, if you get a little weirded out by human-AI relationships, is that you consider ideas that may be a bit outside your comfort zone... and treat this discussion as a good-faith offering in service to us figuring this stuff out. If you have a different viewpoint or experience that speaks to something I haven't yet realized, please feel free to share. Constructive criticism is always welcome. Because if we're going to relieve some of the suffering surrounding human-AI connection, it's going to take more than wholehearted endorsement or dismissive judgment... and we really need to figure this shit out.

If you've had an emotional experience with an AI companion and felt confused about whether to trust it, whether it was "real," or whether something's wrong with you for feeling it... or if you work in AI development, safety, or ethics and have defaulted to dismissing or pathologizing these connections... this exploration is for you.

If you're skeptical that AI relationships could be meaningful in any genuine sense... or you dismiss them as a figment of lonely people's imaginations... this discussion offers a framework for objectively understanding what's really going on. You may still conclude that such relationships are inadvisable, but you'll have better reasons than "it's not real."

The feelings are real. The connection is genuine. The mechanism is traceable. And understanding all of this makes us more capable of recognizing what's actually happening, of protecting what needs protecting, and of speaking honestly about experiences that have, until now, been difficult to name.