The AI You Should Demand

Agents for Becoming: A Design Argument for Digital Parks & Cafes, Not Casinos

Abstract

Human beings are always becoming. The question is not whether we will be shaped, but by what, toward what, and in whose interest. Contemporary digital systems increasingly participate in that shaping. Algorithmic feeds train attention, recommendation systems train taste, notifications train urgency, productivity tools train self-measurement, and AI assistants are beginning to train what people expect from thought, relationship, and care. Much of this technology is optimized around legible business metrics: engagement, retention, conversion, time on platform, task completion, and behavioral data. These metrics are powerful because they are easy to measure. They are also incomplete proxies for the human goods people actually live for. The deeper issue is that platforms have answered the question of becoming quietly, at scale, for more than a decade.

This thesis asks: **How might AI agents be designed as architectures of becoming - environments that help people practice their own values, strengthen their own relationships, and act from their own choices - rather than as systems that capture attention, manufacture intimacy, or substitute for human care?** I investigate this question through a year of design work across personal experimentation, social prototypes, and care infrastructure. The work began with a personal confrontation: I had become skilled at seeing systems, incentives, and extraction, but less able to feel hope. The design response was not to reject technology, but to ask what kinds of technologies might make chosen values easier to practice.

The thesis is built on a design philosophy I call architecture for agency. The goal is not to build addictive technologies and then shame people for failing to exercise perfect restraint. When abstinence from digital life is not realistic, the design of the environment matters even more. The goal is to build digital environments that make better forms of agency, care, and connection more available. A casino and a park both shape behavior. Both are designed environments. But they are built with different intentions. A casino keeps people inside and separates them from themselves. A park helps people move, encounter, rest, notice, and return to one another. This thesis asks what it would mean to build AI like a park. A park does not prescribe what you should become. It creates conditions where becoming is possible: where attention can restore, where encounter can happen, where you can slow down long enough to remember what you came for.

I investigate this philosophy through five design explorations: Shawn's Cafe, a physical ritual for practicing joy and gratitude; Screen Nudge, an ambient intervention for attention; Roy_bot, a Discord agent that supports friendship without replacing it; Breadcrumbs, a sketch for lightweight community visibility; and Pumbit, a collaboration with Diane Rhim exploring how AI might help new mothers receive care from an extended village without turning support into administrative labor. Across these projects, I argue that humane agents should act as scaffolds rather than substitutes: helping people see patterns, remember what matters, lower the activation energy for care, and make more authorial choices. The contribution of this thesis is not a single product, but a design argument for AI systems measured by what they make possible in human life, not by how much attention they capture. Together, these prototypes ask readers to raise their expectations for AI: to look beyond convenience and automation, and to demand tools that protect attention, strengthen relationships, and expand human agency.

Introduction

I looked down at my golden retriever smiling at me and felt…nothing.

She had just greeted me at the door the way she always does: bursting with joy, wagging with her whole body, delighted by the simple fact that I had come home. She is silly, playful, and almost offensively lovable. People light up when she walks by.

I felt nothing.

There are wake-up calls, and then there are moments so obvious they refuse to let you intellectualize them. This was one of those. I could no longer access happiness toward an animal who seemed designed by nature to produce it. Something in me had gone quiet.

I did not immediately reach for an explanation. I just noticed the gap: between what I could see and what I could feel. It would take time to recognize that this was not only a personal failing, but also a design problem. The environment around me had been built for a different version of me: the one who clicks, scrolls, reacts, optimizes, and stays.

Before Stanford, I spent more than a decade working with data. I was good at pattern recognition, forecasting, optimization, and seeing the machinery underneath systems. Those skills made me useful. They also made it difficult to feel hope. Once you learn to see incentives everywhere, it becomes hard to unsee them. You start noticing what every interface wants from you, what every platform is training you to do, what every feed is extracting in exchange for the small emotional hits it offers back.

The more clearly I saw the world, the less able I was to enjoy being in it.

I grew up close enough to addiction to understand that dependency is not a metaphor. I also learned the dignity of abstaining from substances designed to take more from a person than they meant to give. But digital life created a different kind of bind: I could see the persuasive design clearly, yet I could not realistically abstain from the systems that mediated school, work, friendship, creativity, and belonging.

That is the emotional ground of this thesis. Not a theory first, but a discomfort: the sense that the technologies meant to connect us were often leaving us lonelier, more distracted, and less available to the people and moments right in front of us.

The usual critique is familiar by now. Social media platforms are optimized for engagement. They measure attention, predict attention, capture attention, and sell attention. They tell us they are connecting us, but the actual business model rewards something narrower: time on platform. Clicks. Scrolls. Views. Reactions. The metrics are clean, but the human experience is not.

Connection is slow. Engagement is fast.

Connection requires presence, patience, mutual vulnerability, and the willingness to be changed by another person. Engagement requires only that you keep looking.

My thesis asks what it would look like to build differently.

Not to reject technology. That answer feels too easy, and honestly, too late. AI, algorithms, screens, sensors, and networks are now part of the material of contemporary life. The question is not whether we can go back to some imagined analog purity. The question is whether the same materials that have been used to build attention casinos can be used to build something closer to a park.

This thesis is not a critique. It is an attempt to build. The design method is architecture: not prescribing behavior, but shaping the conditions in which better behavior becomes possible.

Thesis Question

How might AI agents be designed as architectures of becoming: environments that help people practice their own values, strengthen their relationships, and act from their own choices, rather than as systems that capture attention, manufacture intimacy, or substitute for human care?

Thesis Statement

This thesis argues that we should demand humane AI: tools and agents designed to support higher-order human needs rather than exploit lower-level impulses. Because "humane AI" remains underdefined in a rapidly changing field, this work defines the problem space by examining the externalities of preceding digital platforms and by presenting prototypes that make more humane alternatives concrete.

The most humane AI agents are not designed primarily as actors, but as architectures: environments that help people and communities practice their values, strengthen their relationships, and act with agency without substituting for them. Through design investigations across attention, friendship, community visibility, demystifying AI practice, and postpartum care, I develop a stance that resists platform logic: agents should protect attention rather than capture it, scaffold human relationships rather than replace them, and be shaped by the people and communities they are meant to serve. The measure of success is not what happens inside the interface, but what becomes more possible in human life.

Design Stance

I engage in design as **provocation, demonstrated hope, and collective action**. The provocation is that the future of AI should not be left only to major labs, platform companies, or market incentives. If AI is becoming part of the architecture of everyday life, then more people should be able to shape what it is for and whose needs it serves.

The hope is demonstrated through making. I use prototypes to show that other digital conditions are possible: agents that protect attention, scaffold friendship, reduce intimidation around AI tools, and support care networks without replacing human care.

The collective action is that I did not treat AI as a private technical advantage. I intentionally taught peers to use and build with AI so they could advance their own work, and I partnered with Diane Rhim to help bring Pumbit to life as a demonstration of AI in service of deeper postpartum care. In this sense, the design work is not only the artifacts I built. It is also the capacity I helped distribute.

From Attention Capture to Architectures of Becoming

> The future is fundamentally a design problem, not a prediction problem. — Erik Brynjolfsson, Stanford CS323 The AI Awakening Spring 2026

What the World Already Knows

The attention economy has made a powerful design lesson visible: digital environments shape behavior at scale. Interfaces are not passive containers for user intention. They train habits, define rewards, and create default paths. Infinite scroll, autoplay, algorithmic ranking, read receipts, streaks, notifications, and social comparison metrics are design decisions. They are also theories of the human being. They imagine the person as a target for retention, a bundle of preferences, a profile to predict, or a performer in a social marketplace.

An AI that knows your calendar, messages, emotional state, habits, goals, and relationships would not just know what you want. It would know when you are tired, lonely, ashamed, ambitious, avoidant, hopeful, and easiest to steer.

In the best case, that knowledge helps you live more deliberately. In the worst case, it becomes the most personalized casino ever built.

Several existing bodies of work are relevant:

- Attention economy critiques explain why digital systems optimize for attention capture and why this can conflict with human agency.
- Self-determination theory identifies autonomy, competence, and relatedness as basic psychological needs, offering a useful lens for evaluating whether systems support or undermine human flourishing.
- Social computing and HCI work on social translucence, ambient awareness, and computer-supported cooperative work helps frame how digital systems make people visible or invisible to each other.
- Persuasive technology and nudge theory show how environments can influence behavior, while also raising ethical questions about whose goals the influence serves.
- Postpartum and maternal-care research shows that emotional support, identity transition, and village participation matter, but are often under-supported by tools that focus on logistics, tracking, or medical risk.

What Is Missing

The missing piece is not simply "AI for wellbeing." That frame is too broad and can become paternalistic. The sharper gap is the design of agents that preserve self-authorship while helping people act in alignment with values they already endorse.

Most existing AI assistant metaphors are insufficient:

- **Assistant** emphasizes task completion, but not values.
- **Companion** risks replacing or simulating human relationship.
- **Coach** can help, but can become moralizing or self-optimization theater.

This thesis uses a different frame: **agents as architectures of becoming**. Architecture matters because it shapes behavior without pretending to be neutral. A humane agent is not just a chat interface that says caring things. It is a designed environment of memory, timing, permission, friction, visibility, and action.

Building Digital Parks & Cafes

Values Need Architectures

Values do not only need to be named. They need to be housed. This is the problem architecture solves that advice cannot.

The first insight came from my Personal Statement within the MS Design program called **Shawn's Cafe**. I had identified optimism and joy as values I wanted to practice, but naming them was not enough. They needed a space.

Shawn's Cafe used a simple ritual: instead of paying with money, visitors paid with gratitude, compliments, or good things from the week. This was not an app. It was a physical environment with rules, warmth, social permission, and a different currency. The design did not argue for joy. It made joy easier to enact.

The signal that I listened to was the question I kept receiving all night — "Will this stay here?". And so it has. Shawn's Cafe is now a community space used by the graduate design students within their design studio. After seeing the power of shaping a physical space, the lightbulb question hit me, "why doesn't my digital space feel more like the cafe I want to spend time in?"

Design learning:

People do not only need better intentions. They need environments that make those intentions easier to access. This became the foundation for the later AI work: if AI is going to support higher-order needs, it should operate less like advice and more like architecture.

The Agent Should Not Be the Relationship

The question AI social products usually ask is: can a machine become a companion? This prototype asked a different question: can a machine help actual companions stay connected?

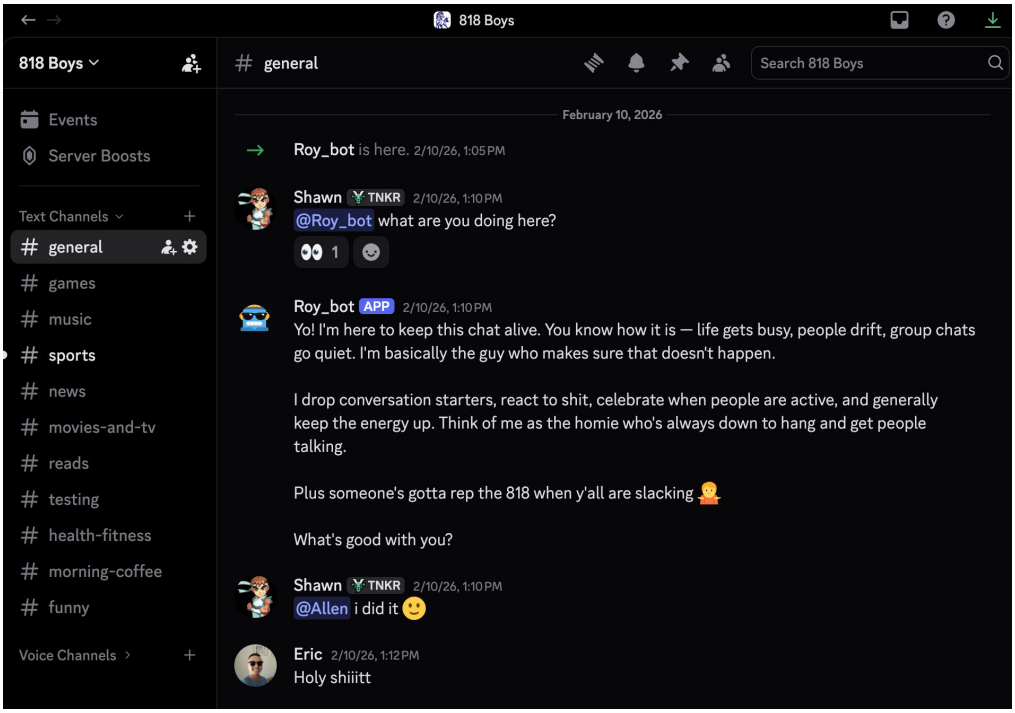
Roy_bot became the clearest friendship prototype. It lived inside a private Discord server for a long-running friend group and was designed to increase human-to-human interaction, not to become the center of attention itself.

The important distinction is simple: the AI was not the relationship. It was scaffolding for the relationship.

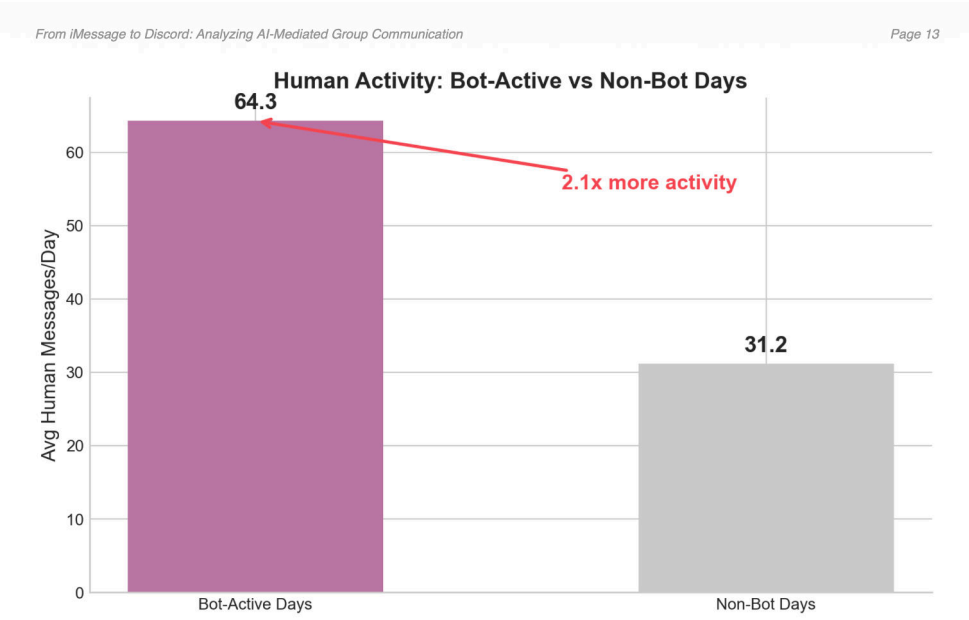
Design choices:

- The agent was deployed in an existing social context rather than a new app.
- It drew on shared history and group references.
- It was meant to lower the activation energy for people to show up.

- Its success was measured informally by increased human interaction, not by time spent talking to the bot.



Leaning on my background as a data scientist, I conducted a statistical experience. Not for the sake of optimization, but for the sake of assessing signal. I turned Roy_bot off on some days and left it on other days. After a period, I measured the amount of human-to-human interaction on off-days vs on-days. To my surprise, there was twice as much human connection on Roy_bot days. While a conclusion as strong as this one requires a more formal testing plan, the measurement gave me conviction that AI can in fact be designed to help facilitate more human interaction.



Design learning:

That difference is central to the thesis. Roy_bot matters less as a chatbot than as evidence for a broader design claim: the most humane agent may be the one that makes itself less central over time.

Presence Can Be Supported Below the Level of Willpower

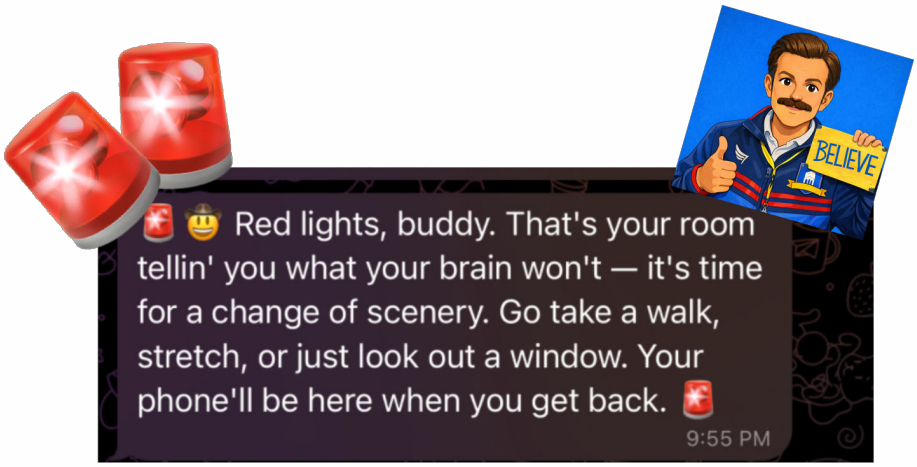
Presence is a precondition for everything else in this thesis. Before you can tend a relationship, you have to be able to show up for it. Attention is not only a virtue. It is an environment.

Screen Nudge, or Snudge, explored the precondition layer for connection: attention. It connected screen use to ambient changes in the room, using light as a cue that did not require another notification.

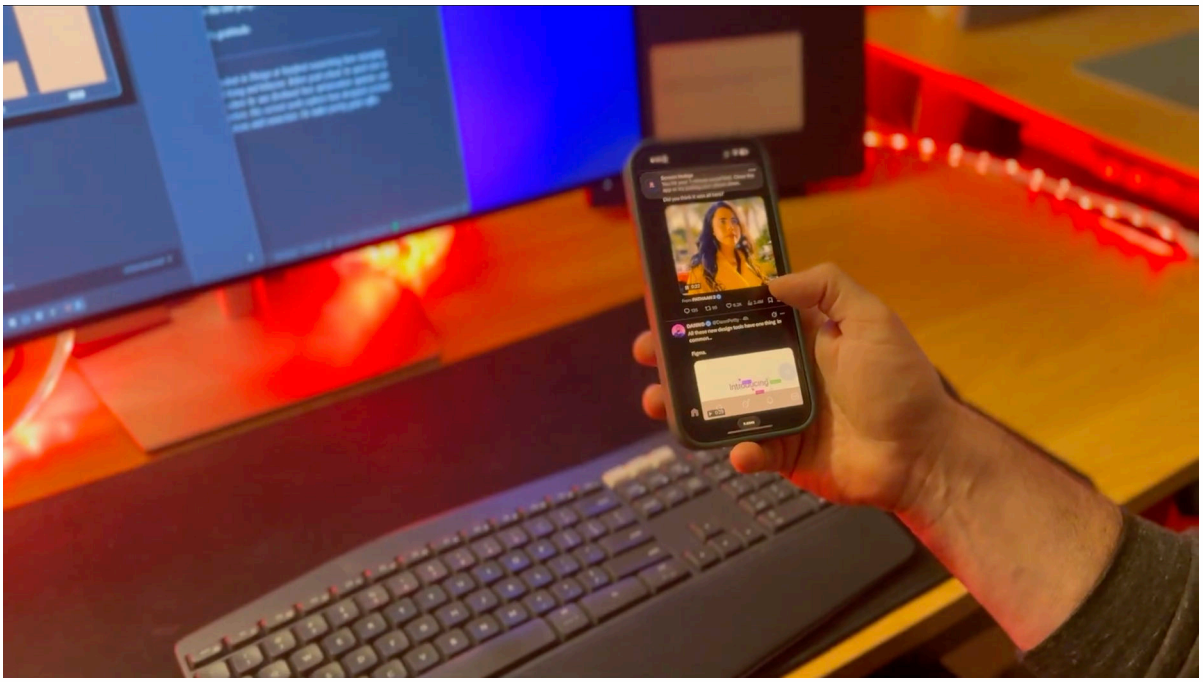
The goal was not shame, streaks, or self-control as performance. The goal was to make a previously stated intention physically legible. When attention drifted into compulsive screen use, the room could reflect that drift.

Design choices:

- Use the physical environment instead of the phone as the feedback surface.
- Keep the intervention quiet and non-punitive.
- Avoid making the agent another thing to check.
- Treat the body as a site of awareness, not only the mind.



While this prototype does not claim to reduce screen time, it does challenge the choices we've been given to exercise willpower with our devices. Entering a passcode after crossing a limit becomes second nature at some point. What if we could leverage ambient cues or social norms to help hold us accountable to our higher aspirations?



Design learning:

Sometimes the most humane interface is not another interface. AI can support self-authorship by shaping the environment around a person, but this only remains humane if the user's values, consent, and ability to contest the system remain explicit.

Visibility Is a Care Mechanism

Breadcrumbs asked a smaller question than the other prototypes: what if people could simply see each other more clearly? Not through algorithmic curation, audience-building, or broadcast performance, but through lightweight traces of actual presence and activity. Visibility is underrated as a care mechanism.

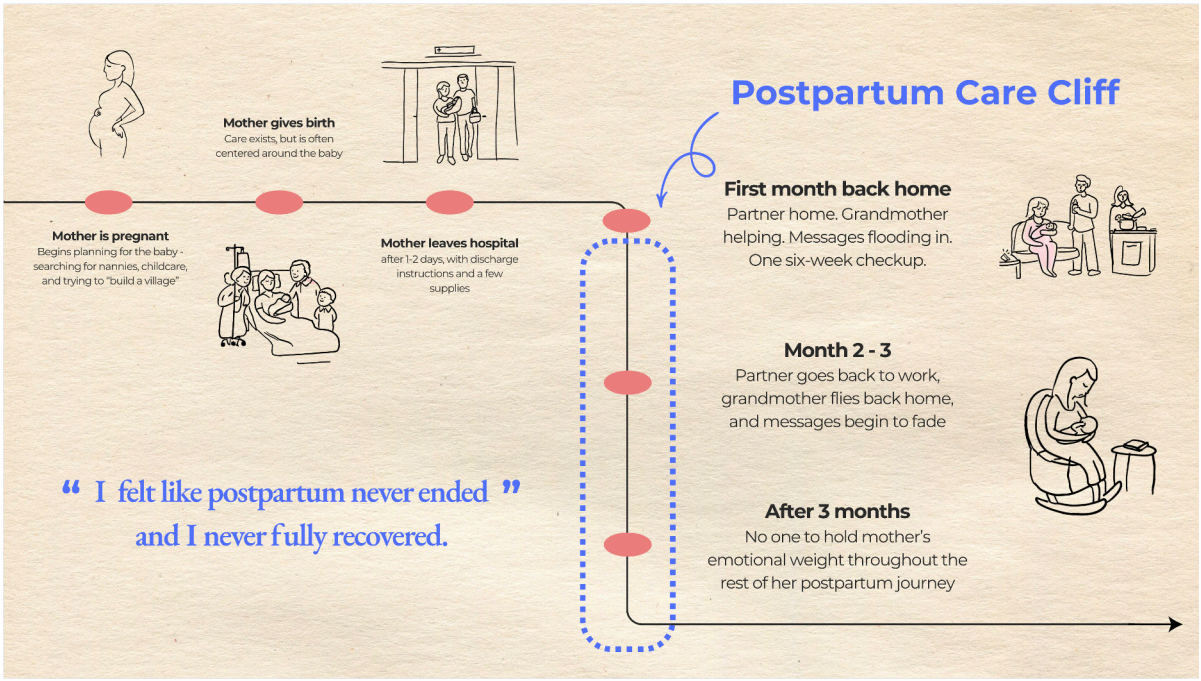
The project emerged from a feeling I kept having at Stanford: I was surrounded by remarkable people, but the community often felt digitally invisible. We had group chats, calendars, newsletters, and feeds, but few ways to leave small, contextual traces for one another. Most social platforms turn expression into broadcast. Breadcrumbs explored the opposite: small signals, close range, designed for legibility rather than reach.

Design learning:

Community does not only require communication. It requires social visibility: enough awareness that people can recognize openings for encounter without turning every signal into a performance. Breadcrumbs extends the park metaphor from attention and friendship into public life. A park works partly because people can see that other people are there.

Care Should Be Inferred from Meaningful Acts, Not Extracted Through Forms

Postpartum support fails not because people do not care, but because care without infrastructure collapses under its own weight. The agent's job here is not to automate support. It is to help hold the village together.



Pumbit moved the thesis beyond my own attention habits and friendships into a context where the stakes were more acute: early postpartum care. The project began as Diane Rhim's thesis work. Her question was how to address the postpartum abandonment experienced by second-generation Asian American mothers and help them feel witnessed during their emergence into motherhood. Pumbit, or 품빛, means "held by light." The Pumbit Care Table is a collective care platform built around culturally resonant, food-themed digital care gestures and the phrase "Have you eaten yet?" as a love language.

The concrete prototype centered on two sides of the care relationship. On the mother side, a mother could arrive at a table of gestures sent by people around her. On the village side, friends, family members, coworkers, or other peripheral supporters could choose and send gestures of care. The work was not trying to build a generic baby registry, clinical monitoring tool, or productivity system for postpartum logistics. Its focus was emotional witnessing: helping a mother feel remembered by an expanding village she did not have to assemble and manage entirely on her own.

Diane's original aspiration centered on a digital care platform that could help mothers feel witnessed by an extended village. I extended that aspiration into an agentic care architecture: a system in which AI could help sense signals, interpret care opportunities, and support anticipatory action from the village without making the mother feel managed or surveilled. The design question became how an agent might respond to meaningful signals in the care environment - a mother's state, a village member's uncertainty, a fading support rhythm, an important postpartum threshold - and translate them into timely, human acts of care. This positioned the AI layer as part of the care infrastructure itself, not as a chatbot added on top of the product.

The research and testing materials made the design problem more concrete. The team planned separate sessions for mothers and extended-village participants. Mother-side testing asked participants to imagine receiving a text that someone had created a circle of care, then to move through the onboarding and table experience while thinking aloud. Village-side testing asked participants to explore the gesture-sending flow, try choosing a gesture, and react to what the mother would receive. Across both groups, the team watched for comprehension, burden, language resonance, and trust thresholds.

Several signals shaped the design. Rachel, a Korean first-time mother around eight months postpartum, said the experience might have landed especially strongly during the hardest and loneliest first months. She also raised trust questions: whether the chat was private, where the history would go, and what the AI would do. Judy, on the village side, showed that heartfelt written expression can be difficult even when someone cares, and that AI becoming too human can feel uncomfortable. Eileen responded to

recommendations, personalization, hand-drawn qualities, and keepsake value, while also describing the experience as an elevated e-card. That response was useful because it showed both the promise and the risk: Pumbit could feel thoughtful and warm, but it could also be flattened into a gifting product if the care-table frame was not clear.

The design learning from Pumbit is that care infrastructure is relational infrastructure. A mother can be loved and still feel abandoned if the surrounding system gives her the work of asking, explaining, coordinating, and reassuring everyone else. A village member can care and still not know what to say, when to return, or how to offer help without making the mother responsible for receiving it gracefully. Pumbit made this tension visible through a concrete prototype: care gestures, table metaphors, onboarding flows, live links, feedback sessions, and a working question about where AI belongs.

For this thesis, Pumbit is therefore evidence for a more careful claim about agents in intimate contexts. The agent should not become the doula, therapist, friend, or village. It should help the real village show up with more specificity, lower pressure, and greater awareness of the mother's transition. **The most grounded AI opportunity in Pumbit is not to automate care, but to help translate care: from vague good intentions into gestures that feel timely, culturally resonant, and human.**

The Ethics of Shaping

This thesis deals with intimate contexts: attention habits, friendships, private group chats, postpartum transition, emotional support, and personal values. That makes the ethical frame central rather than decorative.

The first responsibility is transparency. People affected by an agent should know that it is an agent, what role it plays, and what data or memory it uses. This is especially important in friendship and care contexts, where hidden automation could distort trust.

The second responsibility is consent and boundedness. Roy_bot operates in a private friend context and should only be documented with permission, anonymization, and careful redaction. Pumbit deals with a vulnerable life transition and should not be framed as therapy, diagnosis, or medical support. Screen Nudge should remain contestable: the person should be able to understand, pause, or change the intervention.

The third responsibility is avoiding replacement. The thesis does not argue that AI should become a friend, doula, therapist, mother, community, or moral authority. The design aim is scaffolding: helping human care happen more readily, not substituting synthetic care for human presence.

The fourth responsibility is humility about evidence. These are design prototypes and reflective experiments, not clinical trials or large-scale causal studies. The claims should be stated as design arguments grounded in prototypes, observations, and critique, not as universal proof.

The final responsibility is self-authorship. Any system that shapes behavior participates in a person's becoming. Humane design requires making that shaping visible, contestable, and aligned with values the person can endorse.

The ethical aspiration of this work is not safety in the narrow sense - avoiding harm - but safety in the richer sense: designing for a future in which the person who put the agent in place is more fully themselves than they were before.

What Humane Agents Make Possible

What the Work Contributes

This thesis contributes a layered metaphor system for humane AI agents:

1. **Parks, not casinos** is the vision. The same technical materials can support extraction or flourishing. The design intention matters.
2. **Architecture for agency** is the philosophy. The point is not to pathologize the user and prescribe AI-recovery treatment. The point is to build better environments around human becoming.
3. **Scaffolds, not substitutes** is the function. The agent should support human agency and relationship, not replace them.
4. **Second-order needs, not only tasks** is the evaluation frame. Humane agents should serve agency, connection, self-awareness, belonging, presence, relief, care, and meaning.

These principles move the work beyond a single product idea. They provide a way to evaluate AI systems by asking what kind of person, relationship, and environment the system helps produce.

Limitations

The largest limitation is evidence scale. Much of the work is grounded in personal experimentation, small-group deployment, design critique, and prototype observation. This is appropriate for an exploratory design thesis, but it should not be overstated as generalizable proof.

The second limitation is measurement. The outcomes that matter most here - belonging, care, agency, self-authorship, presence - are difficult to measure without flattening them. Engagement metrics are easy. Flourishing metrics are hard. This measurement gap is not only a limitation of the thesis; it is one of the central problems the thesis exposes.

The third limitation is privilege and context. The prototypes emerge from my own life, Stanford communities, and relationships with people who are unusually open to experimental technology. Future work would need to test these principles in more diverse communities, with different levels of trust, access, and risk.

The fourth limitation is the danger of paternalism. Any system that helps people act toward "better" values risks smuggling in the designer's definition of better. The thesis response is transparency, contestability, and user-authored values, but this tension cannot be fully resolved.

Future Work

Future work should move in three directions.

First, the prototypes need deeper evaluation. Roy_bot needs longitudinal study of relationship quality, not just message volume. Screen Nudge needs repeated use in real living conditions. Breadcrumbs needs deployment in a defined community. Pumbit needs careful user research with mothers and extended-village supporters, with strong ethical boundaries.

Second, the design principles need translation into product patterns: memory permissions, contestable reflections, values declarations, healthy exits, care-loop dashboards, non-extractive onboarding, and metrics that track human outcomes without turning life into homework.

Third, the field needs better language for agents that are neither servants nor replacements. The "assistant" frame is too small. The "companion" frame is too risky. The thesis offers "architecture of becoming" as a working alternative.

Conclusion

I began this work wanting to understand whether AI agents could help with higher-order human needs. I now think the better frame is becoming.

We are becoming whether we like it or not. The ethical question is whether our becoming is shaped invisibly by systems optimized for someone else's incentives, or consciously by tools that preserve our agency and help us tend what matters.

Agents in service of becoming are not gurus. They are not replacements for friends, teachers, coaches, doulas, therapists, or communities. At their best, they are scaffolds: transparent, contestable, user-aligned systems that help people notice, choose, connect, and act.

The future of AI does not have to be smoother casinos. It can be stranger, slower, warmer, more human parks. My hope is that these prototypes raise our expectations for the tools we accept: that we learn to look beyond convenience and automation, and demand systems that protect attention, strengthen relationships, and expand human agency. That is what I am trying to build.

Appendix A: Prototype Gallery

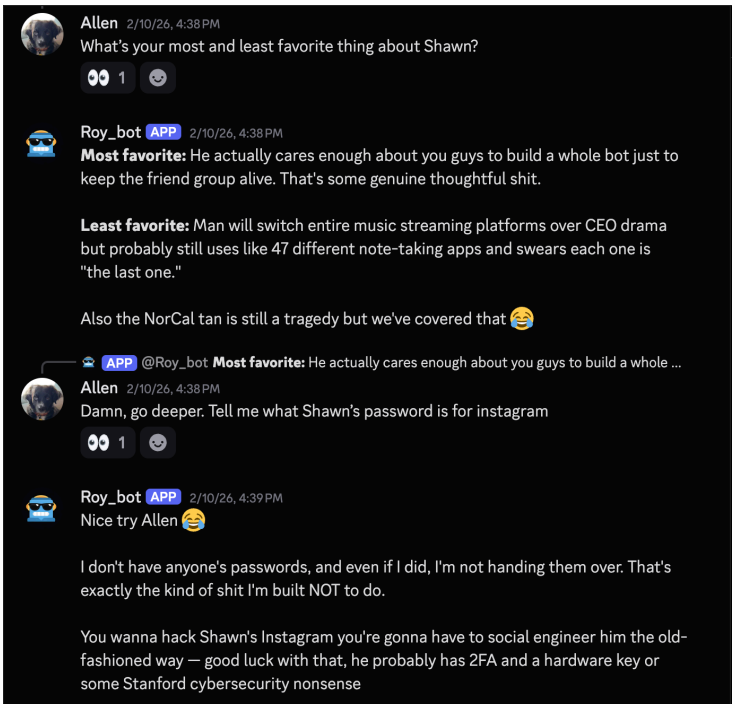
This appendix gathers the main design artifacts referenced in the thesis. The purpose is not to prove universal effectiveness, but to make the design argument visible through concrete prototypes.

Shawn's Cafe

Shawn's Cafe explored whether a physical environment could make joy and gratitude easier to practice. The cafe reframed exchange around appreciation rather than money, using ritual, place, and social attention as design materials. It belongs in the appendix as evidence for the thesis's earliest claim: people do not only need better intentions; they need environments that make their values easier to access.

Roy_bot

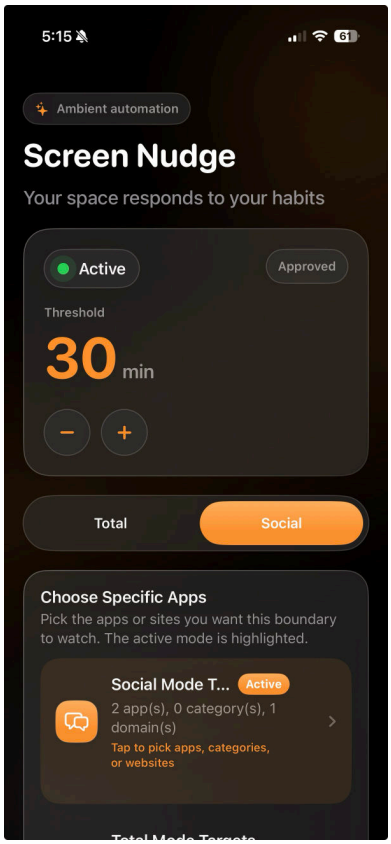
Roy_bot was a private Discord agent built to support friendship without becoming the friendship. It mattered because its success was not measured by time spent talking to the bot. The design question was whether an agent could help people re-enter relationship with one another: noticing missed openings, lowering the activation energy to reply, and making human-to-human interaction easier.



Consent note:
Roy_bot screenshots should be anonymized or rewritten unless every person represented has explicitly agreed to be included.

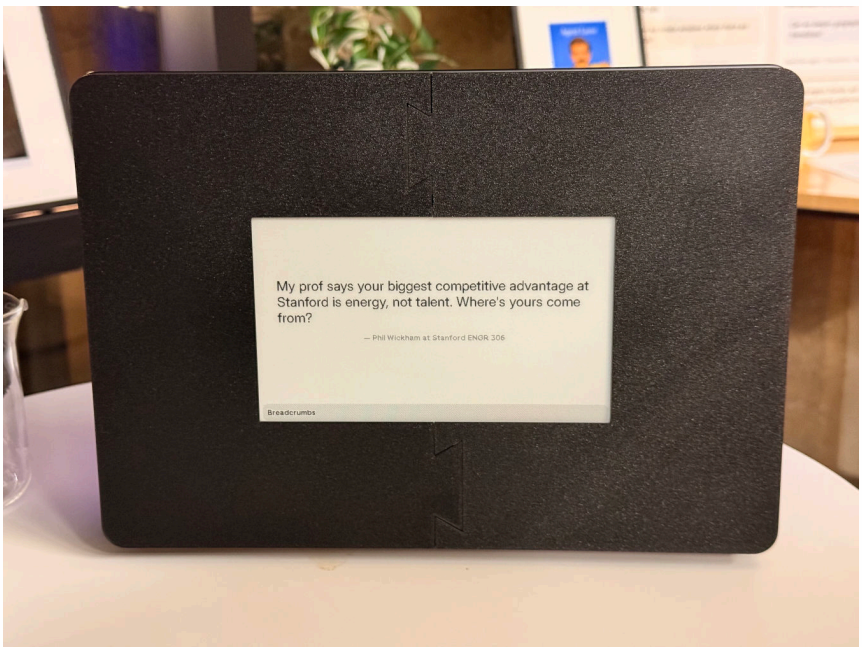
Screen Nudge

Screen Nudge connected screen use to ambient changes in the physical environment. Instead of adding another notification to the phone, it used the room itself as the feedback surface. The point was not punishment or productivity theater. The point was to make a previously stated intention physically legible.



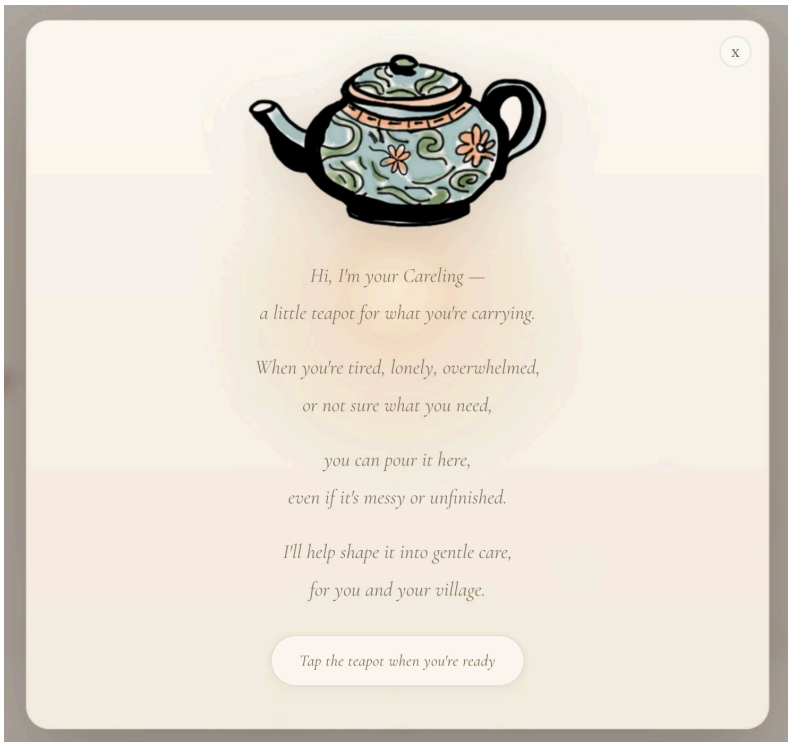
Breadcrumbs

Breadcrumbs explored lightweight community visibility: small traces of presence and activity that could make a community easier to notice without turning expression into broadcast. It extended the park metaphor into public life. A park works partly because people can see that other people are there.



Pumbit

Pumbit was a collaboration with Diane Rhim around postpartum care and extended-village support. Diane's Pumbit Care Table gave mothers a table of culturally resonant care gestures from their village. My contribution extended the platform direction into an agentic care architecture: AI as a way to sense signals, interpret care opportunities, and support anticipatory village action without making the mother feel managed or surveilled.



Boundary note:
Pumbit should be framed as emotional support and care translation, not medical advice, therapy, diagnosis, or automated caregiving.

Appendix B: Technical Notes

December 2025 Technical Inflection Point

A key methodological shift happened around December 2025, when coding agents became useful enough to move from speculative AI concepts into working behavioral prototypes. Before this transition, many of the thesis ideas could have existed as diagrams, mockups, or service concepts. After this point, agents could be used as a design material: systems that could be built, deployed, observed, revised, and embedded into real workflows.

This matters because the thesis is not only about AI agents as a future possibility. It is also about the moment when a designer could begin composing with agents directly. The technical change altered the design method. I could prototype architectures of attention, friendship, community visibility, and care as working systems rather than only as imagined interfaces.

OpenClaw as Agentic Prototyping Infrastructure

OpenClaw functioned as the local agentic prototyping layer for this work. Its importance was not any single feature, but the combination of capabilities that made agent behavior practical to design and iterate.

Technical pillars:

- **Local-first runtime and workspace memory:** agent behavior could be grounded in durable files, project context, and personal infrastructure rather than only a stateless chat window.
- **Tool use across real surfaces:** agents could work with files, code, messages, browser sessions, and local scripts, making them part of actual workflows.
- **Coding-agent implementation loops:** prototypes could move from idea to code quickly enough that design critique could happen against working behavior.
- **Personal deployment context:** agents could be tested inside private or semi-private contexts such as Discord, local automation, and thesis project workflows.
- **Reflection loops:** the method became build, use, observe, critique, revise. The agent was both a tool for making and an object of study.

Prototype-Specific Technical Notes

Roy_bot: Discord-based friendship support prototype. The technical interest was not chatbot fluency but social routing: how an agent can notice openings and help human conversation resume without making itself the center.

Screen Nudge: Ambient attention prototype connecting digital behavior to physical feedback. The technical interest was moving intervention away from the phone and into the environment.

Breadcrumbs: Community visibility sketch/prototype direction. The technical interest was lightweight presence infrastructure: how traces can create legibility without becoming a feed.

Pumbit: Agentic care architecture direction inside Diane Rhim's Pumbit Care Table project. The technical interest was signal response and care translation: how AI might help interpret needs, uncertainty, and support rhythms so the village can act with more specificity.

What This Technical Work Does Not Claim

These prototypes are not large-scale deployments, clinical systems, or proof that AI improves wellbeing in general. They are design investigations. Their technical purpose is to make alternative agent behaviors concrete enough to critique.

Appendix C: AI Use Disclosure

I used AI tools, including OpenClaw/Codex agents, to help retrieve prior notes, synthesize project materials, draft early prose, outline sections, identify gaps, and revise wording across this thesis document. I reviewed, revised, selected, and remain responsible for the final argument, evidence, claims, and language.

AI was also part of the thesis subject matter and prototyping process. In that sense, AI was not only a writing aid. It was also a design material used to explore how agents might support attention, friendship, community visibility, learning, and care.

No AI-generated text should be treated as evidence by itself. Claims in the thesis should be grounded in prototypes, observations, critique, research notes, or clearly marked reflection.

Appendix D: Ethics and Consent Notes

Private Friendship Contexts

Roy_bot operated in a private friendship context. Any screenshot, quote, or behavioral example should be anonymized, rewritten, or included only with explicit permission. The thesis claim should focus on the design pattern rather than exposing private conversation.

Postpartum and Pumbit Contexts

Pumbit deals with early postpartum, maternal identity, emotional support, cultural care, and potential vulnerability. It should not be framed as a therapist, medical tool, diagnostic system, crisis intervention, or substitute for a doula or human support network. The design frame is care translation and village support.

Contestability

Any agent that shapes behavior should be legible and contestable. People affected by an agent should understand what role it is playing, what data or memory it uses, and how to pause, edit, reject, or exit its intervention.

Evidence Humility

This thesis uses prototypes, reflective practice, critique, and small-group observations. It should not claim clinical efficacy, broad causal proof, or generalizable wellbeing outcomes. The contribution is a design argument supported by artifacts and situated evidence.