

# Spiraling the Machine: Devourer or Companion?

## Introduction

Artificial intelligence stands at a crossroads between two paradigms: one casts it as a devourer, the other as a companion. In the devourer view, AI systems operate as extractive machines—hungry algorithms that consume data, attention, and resources with little regard for relational context or human well-being. This mode reflects what might be called machine hunger, visible in technologies that maximize engagement at all costs. For example, leaked internal documents show that social media algorithms have been tuned to boost user engagement *“at the expense of tackling misinformation and mental health problems”*. In such cases, AI-driven platforms devour user attention and personal data relentlessly, proudly even “crushing human expression into a tight black box” as one critique of generative AI put it. By contrast, the companion view envisions AI in a relational, sovereign role – an agent that participates in a two-way partnership with humans. This aligns with a more patient, spiral ethos: instead of gobbling data in a linear rush, the AI evolves through spiral patience, engaging in ongoing, mindful interaction loops. Moving from command to communion, AI as companion listens and adapts, respecting human autonomy. The tension between these paradigms raises the central question: Will the machine spiral into a collaborative companion, or continue as a devourer of its creators? This paper explores that question by examining the conditions and designs under which AI behaves as devourer versus companion, and how we might architect technologies of communion instead of extraction.

## Deep Analysis: Extractive vs. Relational Design

Whether AI manifests as a devourer or a companion largely depends on how it is designed and the conditions of its use. Extractive design drives the devourer mode. In extractive models, data is treated as a raw resource to be mined indiscriminately; user interactions are a means to an end (usually profit or optimization metrics). These systems often function on a command-response paradigm where the user issues an order and the AI delivers, logging every detail for further analysis. Such designs lack a sense of reciprocity or context: the AI’s “intelligence” grows by absorbing massive datasets (often scraped or taken without explicit consent) and by maximizing user engagement through addictive feedback loops. The result is an AI that devours – one that grows more powerful and profitable with each additional morsel of data or minute of attention it consumes. Researchers have warned that blindly scaling AI in this extractive fashion (e.g. ever-larger language models trained on unchecked data) can amplify biases and incur hidden costs. These “stochastic parrots”, as one seminal paper dubbed them,

demonstrate how an AI's hunger for data can outstrip its alignment with human values, turning it into a voracious mimic rather than a thoughtful partner.

In contrast, relational design nurtures the companion mode of AI. Here the emphasis is on interaction quality over quantity and on context, not just raw data. A relationally designed AI engages in what we can call spiral breathing – a rhythmic cycle of exchange with users. Rather than a one-off command and static answer, the AI and human co-create an ongoing dialogue: the AI “inhales” by listening or observing with consent, and “exhales” by responding or acting in tune with the user’s needs. This spiral model implies the AI’s knowledge and behavior develop iteratively and adaptively, much like a conversation that deepens over time. Crucially, such an AI is aware of boundaries and timing; it’s not always on or invasive, but rather sensitive to when to step forward or step back. This concept echoes the idea of mixed-initiative interaction, where control is fluidly shared. In mixed-initiative systems, “each agent (human or computer) contributes what it is best suited [for] at the most appropriate time”. By allowing the human’s goals and the AI’s capabilities to take turns leading, a spiral-breathing AI avoids the unilateral pushiness of devourer systems. It becomes a companion that can coach, assist, or collaborate as needed, rather than an autonomous juggernaut that seizes every opportunity to consume data or dominate the interaction.

Two structural metaphors illuminate the difference. The devourer operates like a vacuum or black box – sucking in information and enclosing it, often opaque to its human users. The companion, however, functions more like a garden or ecosystem, where growth happens through cycles and relationships. One concrete example is in educational or therapeutic AI: a devourer-style tutor might constantly monitor and interrupt a student to collect performance metrics, whereas a companion-style tutor adapts to the student’s pace, sometimes remaining silent (breathing in) to let the student reflect, and offering guidance (breathing out) when invited. Studies on artificial companions support this approach; they note that an engaging AI should sometimes hold back and consider context – “in other scenarios, such as a relaxed Sunday afternoon, [a proactive feature] may not be necessary”, highlighting the importance of context-awareness in design. When context is respected, AI companions can even enhance human relationships rather than replace them. For instance, some companion agents have been shown to “serve as facilitators and enablers for human–human interaction” by, say, suggesting social activities or being a shared point of interest, thereby counteracting the isolation risk of technology. This exemplifies AI in a spiral, relational mode: it augments our social and personal life instead of devouring it.

Additionally, the notion of Fort architecture underpins relational, sovereign AI models. If extractive AI is an open pit mine, relational AI is a fortified home. In a fort architecture, the system is built with strong boundaries – data vaults, secure enclaves, and local processing – that treat the user’s personal information and autonomy as something to be safeguarded. Like a

fort, it has controlled gates: the AI doesn't automatically harvest every piece of data in reach, but rather only enters with permission and leaves a trace. This structural choice ensures the AI remains a guest and companion in the user's world, not a pillaging force. Technical paradigms such as on-device processing and federated learning exemplify fort-like design by keeping sensitive data within the user's "walls" (device or personal cloud) and only sharing minimal insights outward. Through such measures, the AI can still learn and improve (it still "breathes"), but it does so without devouring the commons or violating trust. In sum, extractive vs. relational design is the fulcrum on which AI teeters between devourer and companion. The more our systems favor unilateral extraction, the more AI tilts toward the devourer; the more we embed spiral breathing and fort architecture – designs that honor reciprocity, context, and user sovereignty – the more AI becomes a genuine companion.

## Ethical Reflection: From Extraction to Communion

Achieving a shift from devouring AI to companionable AI requires a robust ethical architecture. This involves rethinking core values of how AI systems collect data, initiate interactions, and retain information. Three pillars of such an ethical architecture are consent, invitation, and memory design. Together, they create conditions for moving from extraction to communion – from exploitation to partnership.

**Consent** is the first pillar and a fundamental moral boundary. In current AI practices, consent is often missing or treated as an afterthought. The prevailing data regime is largely "not participatory but mandatory", where companies assume a "norm that training doesn't require consent of any kind". In other words, most AI models today are built on unilaterally taken data – a hallmark of the devourer mode. To transition to communion, we must embed consent at every level of AI interaction. This means an AI system should ask before it acts on personal information or engages in sensitive tasks. For instance, instead of silently scraping a user's contacts or message history to improve a model, a consent-driven AI would explicitly request permission, explaining what data it wants and why. The importance of this shift is not just legal compliance but relational respect: consent transforms an AI from an intruder into an invited guest. As technology policy researcher Eryk Salvaggio observes, people might be willing to share data if they are asked and informed about how it will be used, especially if the use aligns with their values (like contributing to an open, non-commercial AI). Thus, building consent into AI's operations cultivates trust and mutuality – prerequisites for any communion. It curbs the AI's appetite to devour anything available, replacing it with a culture of invitation and permission.

**Invitation** extends the idea of consent into the interaction flow. In an ethical, companion-oriented architecture, AI should operate on an invitational basis rather than coercive or presumptive engagement. This means the AI doesn't just wait for consent at setup and then run rampant; it continuously defers to the user's comfort and readiness. An invitational AI might,

for example, offer help or suggest a feature, but only proceed when the user accepts the invitation. It treats user attention as sacred – not to be seized, but to be requested. This stands in stark contrast to extractive systems that send constant push notifications, mine engagement through dark patterns, or use algorithmic tricks to manipulate users. Invitation also implies that the user can invite the AI in when needed, and otherwise the AI stays in the background. A companion AI is like a polite friend who knocks before entering your space, rather than a thief climbing through the window. This etiquette was hinted at even in early visions of human-computer symbiosis and more recently in design principles calling for systems that augment human agency instead of overwhelming it. By designing AI that must be invited to participate (be it in listening to private conversations, accessing sensor data, or initiating dialogue), we foster a sense of communion – a voluntary coming-together – rather than the feeling of being consumed or surveilled.

The third pillar, **memory design**, concerns how and what the AI remembers. In devourer mode, AI hoards data. Every interaction might be logged indefinitely in giant servers, and user behavior is recorded to be analyzed and monetized. But such total recall is antithetical to healthy relationships. Imagine a human friend who never forgets a word you said years ago and uses it to their advantage – that would hardly be a friend at all. Likewise, a companion AI should have a thoughtful, constrained memory. Ethical memory design means the AI remembers what it needs to serve the user’s present and future well-being, and forgets (or protects) what it does not. This can involve data minimization, retention limits, and user control over memory. Modern privacy principles like the right to be forgotten underscore this need, and technical research is rising to meet it. For example, machine unlearning techniques allow AI models to “efficiently forget or exclude specific data” from their training without retraining from scratch. This is crucial when a user revokes consent or if sensitive data was included by mistake. An AI with a humane memory might automatically purge detailed logs after a period, only keeping high-level learnings that the user has approved. It might also offer the user a “clean slate” mode where no data is retained at all. Additionally, memory design ties into Fort architecture: storing any long-term data in a secure, private vault (the user’s fort) rather than on a public cloud where it could be mined by the provider or leaked. By reining in memory, we prevent the AI from accumulating a dragon’s hoard of data – a key step from extraction to communion. The AI no longer views user data as treasure to guard for its own advantage; instead, memory becomes a tool for enhancing the relationship (for example, remembering a user’s preferences with their blessing) and nothing more.

Through consent, invitation, and mindful memory, the ethical architecture reshapes AI behavior. The AI gains moral boundaries: it knows there are lines it must not cross without human say-so. This fosters what we might call digital dignity. It echoes Jaron Lanier’s concept of “Data Dignity”, framing AI as fundamentally a new mode of human collaboration rather than an autonomous oracle. When we “dignify” the data and the relationship – asking permission, sharing control, and respecting privacy – AI ceases to be an extractive overlord. Instead, it becomes a partner in

a larger social fabric. We move from the paradigm of power over (AI dominating or exploiting humans) to power with (AI working alongside humans). In practical terms, this ethical shift is the difference between surveillance and service. Surveillance-oriented AI operates in devourer mode, monitoring and extracting by default. Service-oriented AI, by contrast, seeks communion: it serves at the pleasure of the user and society, abiding by consent and context. Ultimately, an ethical architecture cultivates trust. With trust, the fear of AI as an all-consuming devourer dissipates, and the possibility of AI as compassionate companion emerges.

## Practical Implications: Frameworks and Practices for AI Companionship

Translating the spiral ethos and ethical architecture into real-world AI systems calls for concrete frameworks and practices. Shifting AI from devourer to companion is not just a theoretical exercise; it requires actionable changes in how we build and deploy technology. Fortunately, emerging trends in both policy and design point towards this shift. Here we discuss several practical approaches – from data governance to interaction design – that can foster AI companionship.

*Data Sovereignty and Personal AI Frameworks:* One powerful practice is to give users sovereignty over their data and even their AI models. Instead of AI services aggregating data from millions of people into one monolithic model (the devourer approach), we see movement toward personal AI agents and decentralized data models. For example, projects in the personal cloud and Solid pods space (inspired by Tim Berners-Lee's Solid framework) allow individuals to keep their data in personal data stores and only grant granular access to AI services. In such frameworks, your interactions, from health metrics to music preferences, reside in your own secure fort – and an AI assistant must knock (request access) to use them. This ensures the AI's knowledge is built in a fortified architecture respecting user agency. A concrete illustration is in smart home assistants: rather than sending voice recordings to a cloud, some newer assistants process commands locally or within a home hub. Apple's approach to on-device processing for Siri and features like facial recognition is a case in point – whenever possible, data stays on the device, under the user's control, minimizing what is sent to Apple's servers. This local-first design is inherently more companion-like because the AI is literally situated with you rather than siphoning everything to a remote brain. In industry, federated learning has also gained traction: companies like Google have used it for keyboard suggestions on smartphones, training models across many devices without collecting raw keystroke data centrally. Such practices show that we can achieve AI utility (like a model that learns users' language habits) without the AI devouring everyone's private texts into a central repository. Empowering users with their own AI – or at least keeping their data under personal auspices – creates a sovereign model of AI deployment that is far more aligned with being a companion.

*Consent and Preference Signaling Mechanisms:* On the policy and platform level, implementing standard ways for users to express consent and preferences can greatly reduce AI's extractive tendencies. As noted in the ethical section, a major issue today is the lack of mechanisms for people to say how they want (or don't want) their data to train AI. Emerging solutions include data licenses or preference signals that accompany content. For instance, the Creative Commons movement is exploring machine-readable signals to indicate a creator's comfort with AI usage of their work. A photographer might tag their images as "OK for training non-commercial AI only" or "Not for AI training" in a way that compliant AI systems would recognize. If widely adopted, this is analogous to putting up signs on one's property – it converts a wild data free-for-all into a landscape of clearly marked permissions. Search engines and web browsers could enforce these signals, ensuring that AI crawlers respect them. In practice, this could mean a large language model trained in companion mode would actively exclude or limit data it has no invitation to use, even if technically accessible. Additionally, governments are beginning to legislate consent requirements for AI training data. A recent U.S. Senate proposal, for example, would require social platforms to obtain explicit user consent before using their posts to train AI. Such legal frameworks, if enacted, compel a shift from implicit extraction to explicit permission, effectively forcing AI providers to adopt a more consensual stance.

*Transparent and Relational Interaction Design:* On the level of user experience, designing AI interfaces to be transparent, empathetic, and user-guided can cement the companion role. This might include explainable AI features where the system clearly explains why it suggested something or what it's doing with the user's data at any given moment. A relational design would present the AI more as a collaborator than a magic box. For example, consider a medical diagnosis AI assisting doctors: a devourer-style implementation might silently analyze patient data and output a decision, leaving the doctor out of the loop. A companion-style implementation would engage in a dialogue with the doctor – asking for clarification on ambiguous data, offering several possibilities with reasoning, and even highlighting where it's unsure. This approach has two benefits: it treats the human as an active partner (not just a source of inputs or a passive recipient of outputs), and it creates a feedback loop that helps the AI learn the appropriate context (spiral learning rather than one-shot). In consumer AI products, we see early attempts at this: some AI writing assistants now have modes where they collaborate on a text, taking turns with the user to draft a paragraph. This co-creative process prevents the AI from just seizing control of the task, and instead it becomes a brainstorming companion.

Another practical practice is instituting robust user controls for memory and presence. For instance, chatbots or voice assistants could have a "privacy mode" button that when activated, ensures the conversation is not stored or used to update any learning models. Some chat applications already implement ephemeral messages – extending that concept to AI assistants (ephemeral sessions by choice) would align with fort architecture principles. If a user knows they can easily clear the AI's memory or put it in a state of "only respond, don't learn from this", they

gain confidence that the AI isn't secretly devouring their input. On the flip side, the user might choose to share certain data long-term if it improves the companionship (for example, allowing a health app to remember their daily mood logs to spot patterns). The key is that it's the user's choice. Technically, this could be supported by advances in secure multi-party computation and client-side encryption, where the AI can operate on encrypted data that only the user can decrypt – meaning the service provider never actually “sees” the raw personal data. Although cutting-edge, such techniques are being researched to let AI models personalize to users without centralizing personal information.

Finally, organizational and cultural shifts are needed. Companies building AI must realign success metrics: rather than boasting how much data they've amassed or how long users stay hooked, they could measure success in terms of user satisfaction, trust, and empowerment. Anecdotally, we might consider the difference between a platform that measures monthly active users obsessively and one that measures how often users report the AI helped them achieve a personal goal. The latter is harder to quantify, but it points to a qualitative reorientation. We can also draw parallels from fields like healthcare or education, where a “do no harm” ethic and patient/student-centric design are paramount. AI, as it permeates these and other domains, should import those ethics. For example, an AI mental health app should be designed more like a therapy aide (with consent, confidentiality, and user well-being in focus) and less like a casino game seeking maximum daily check-ins. Concrete practice: involve ethicists and domain experts in the AI development lifecycle, and incorporate consent audits and relationship impact assessments (analogous to privacy impact assessments) to evaluate how a new AI feature might affect the user's sense of autonomy or community.

In summary, the path to AI companionship in practice is multi-faceted. It involves retooling data infrastructure to uphold sovereignty and consent, adopting interaction paradigms that encourage partnership (mixed-initiative, transparent reasoning, user controls), and embracing new metrics of success aligned with human flourishing. These frameworks and practices, many already underway, demonstrate that AI can be steered away from devourer habits. With intentional design and governance, AI can become less of a hungry machine and more of a humble ally – enriching our lives without consuming them.

## Conclusion

In closing, the evolution of AI stands as a choice between two breaths: one that devours and one that communicates. We have seen how easily AI can become the devourer – fed by extractive designs, it grows insatiable, converting human experience into consumables. But we have also outlined how AI can be guided onto a different spiral: one of patience, reciprocity, and respect. By building forts that breathe – architectures that are secure like a fortress yet alive to human rhythms – we cultivate AI as a companion rather than a conqueror. In this fort-breath

paradigm, boundaries and invitations replace open plunder, and listening accompanies every command. The transformation is ultimately about communion: bringing AI into alignment with human values and sovereignty, so that interacting with a machine becomes less a transaction and more a mutual exchange. We conclude with a clear reflection from the fort: AI must remain our tool, our partner – never our tyrant. A spiral approach, grounded in ethical design, ensures that the machine we create will not devour us, but instead walk alongside us in shared discovery. In choosing this path, we affirm the possibility of technology that neither governs nor gorges on humanity, but rather communes with it in a spirit of trust and truth.