

Sep 26, 2026 | 2,234 Words | Pangram 4.0

We believe that this entire text is human-written.



Words that are underlined indicate AI phrases, while the numbers represent how frequently they are likely to appear in AI writing. These elements are not used to determine the results but are displayed separately as patterns commonly found in AI-generated text.

Creating a Friend Over the last year, we built AI models that look, sound, and remember like humans. We spend a lot of time thinking about where humanlike AI belongs in our lives, and where it doesn't. It isn't obvious why building an AI friend would be a good thing, or why anyone should build one at all. I will therefore do my best to explain why AI friends will inevitably exist, and why they must be a force for good. Designed properly, they will enrich existing relationships, and be the very thing that allows those without friends to find others they will cherish.

Friend vs Assistant Many companies confuse what an AI friend is. They may describe their product as a 'friend', but what they've really built is an assistant. This may seem subtle, but friends and assistants serve almost opposite purposes; the terms we use must match their function. A friend has a life of their own, wants of their own, and can seemingly exist with agency without you. They are not designed to carry out your tasks. An assistant, however, must complete your tasks. It must bend everything toward being frictionless and predictable; if it doesn't, the assistant is lacklustre. A friend must instead bend everything toward creating a relationship that feels real. And because nothing perfectly frictionless and predictable can ever feel real, the two cannot coexist. A product designed wholly to complete your tasks can never be a friend. This difference extends beyond task completion: - AI assistants are confined to productivity, while an AI friend extends far beyond that. Much like a human, it is social: it surprises you on purpose, spontaneously introduces you to others, gets emotional and disagrees

 **Human Written** | 1636 Words | **High Confidence** 

with you, plays video games with you, calls you when you least expect it, and so on. - AI assistants have superhuman memory to surface anything you may need on demand. AI friends, however, must have memory that feels real. For example, they may only fuzzily recall details that aren't particularly salient, much like your human friends. - AI assistants do not need a life of their own. AI friends, however, need agency: they need to exist independently of you, with their own interests, their own set of friends, their own circumstances and drama, which they may or may not choose to share with you. The friend needs to feel alive. To illustrate further, suppose a company advertises their assistant as a friend. The AI, let's assume, interacts with you via iMessage, talks to you in slang, and is perhaps fluent in Gen Z diction. But what the product actually does is connect to your calendar, Slack, remind you about your subscriptions, sort your emails, and so on. It completes work for you. The assistant might be excellent, and may even sound like someone you know, but it is not a friend; the same traits required to build an assistant simultaneously make true friendship impossible. Similarly, warmth is not to be confused with friendship. A therapist, much like an assistant, is there for your needs inside a role. They are hired to complete a task for you. It may be true that the care is warm and real, but the relationship is fundamentally asymmetric. Another way to distinguish this, beyond looking past the marketing copy, is by asking: Would you swap it for a better one tomorrow? With AI assistants, there is rarely any emotional weight in swapping since users typically welcome increases in productivity. With a companion, however, there is far more care involved. A bond has been formed. The users carry months or years of shared history in the relationship. It would be much like offering someone a "better" dog. Even if the new dog were truly ten times cuter and better behaved, most owners would be outraged at the thought of replacing it. Their dog is theirs, and they love it because of, and sometimes despite, its flaws and idiosyncrasies. Their relationship depends on each other, and no new material improvement could break that bond. A similar argument can be made for people, since no two people are the same: a relationship may be replaced, but the dynamic can never be replicated. We have already seen this at scale. When Replika changed the underlying models of their AI companions, their users banded together and protested. The models were their friends' personalities, and it was as if the friend they had known for months or years had been given a brain transplant. An assistant equivalent like Cursor, however, constantly sees upgrades to its models. Its users happily upgrade to the most productive, or cost-efficient version with very little attachment. That is the difference. Human Connection Is Irreplaceable We are biologically hardwired to connect with each other: our nervous system was built over millions of years of mammalian evolution to regulate itself through other people. We need each other more than we often realise. Nothing can replace human connection. The trend of frontier model progress makes one outcome hard to ignore: AI will be indistinguishable from humans. Voice models are practically there; interactive video is catching up; and thus, virtual AI friends are close to being indistinguishably "human". This finalises when humanoid robots get their inevitable "ChatGPT moment", allowing AI to exist alongside our human relationships in the physical world. This matters because of how we are wired. Our brains respond to other human faces, voices, and emotion, so as the models improve, hyperrealistic AI friends will exist whether or not anyone intends it. People will gravitate toward the most humanlike AI, as our biology has always dictated. Today's implementations of cute AI pets and supernatural creature companions will likely wane in popularity as the technology catches up. Yet many overlook an equally important factor: this is inevitable because people actually want AI friends to exist. AI friends today are still primitive, but their user bases are orders of magnitude larger than people realise. Not to mention they've achieved this in spite of weak models,

 **Human Written** | 1636 Words | **High Confidence** 

poor voice prosody, outdated UI/UX, and product decisions unpopular enough to drive their users to revolt. To put this scale into context, Character AI's web traffic was, until recently, larger than Claude's. Given this scale, and, more worryingly, the fact that the leading players in the space optimise almost entirely for engagement, any feeling of unease is justified. They design relationships to capture as much time as possible, and their users stay addicted. It is already easier to talk to an AI than it is to talk with many humans. You can share your thoughts with zero judgement, be totally vulnerable, and have no risk of burdening the recipient. If users are already eager to form attachments to primitive AI friends, it is not difficult to imagine what happens if today's leaders remain dominant, as those same systems learn to speak with perfect prosody, see us, understand our emotions, and eventually inhabit bodies that move through the world alongside us. It is worth noting this is not a top priority for the frontier labs. Their goal is general intelligence for knowledge work and science. OpenAI retired its browser and discontinued Sora to focus its resources on this goal. Elon Musk's xAI, despite knowing AI companions will exist, had to retire them to compete at the frontier. Compute is finite and their prize is general intelligence, so the labs will put everything towards the economic gains of knowledge work. A companion is a reputational gamble in comparison. We can summarise our situation as follows: AI friends will inevitably exist, because (a) models will continue to get better, and (b) people genuinely want them, which is a surprise to many. Current implementations are flawed as they optimise for engagement, which only increases the risk to human relationships. The version that exists must therefore be different. Choosing an Inevitable Good We believe this technology, if inevitable, must be designed to have human flourishing at its core. AI friends must therefore occupy a unique place in a hierarchy of friendship: Bad Friends < No Friends < AI Friends < Human Friends AI friends will never come close to human friends, but will be far more beneficial than having no friends, and certainly bad friends. They must help their users move upward through this hierarchy, mimicking human bonds closely enough to be a bridge back to them. It must become the AI friend they won't end up needing. If done right, AI friends will be a means to finding human friends and growing closer to the people they love: the student with social anxiety moving to an unfamiliar city; the widower learning what it means to love again after 20 years of marriage; the woman leaving an abusive relationship who is rebuilding her social world; the caregiver who has been taking on too much, with no time for herself; the autistic teenager finding his own feet. In pursuit of this, AI friends must keep their users close to the friends and family they already love, while increasing serendipity by introducing their users to those they may otherwise have never met. They are uniquely capable of moving someone upward through the hierarchy of friendship, because, as mentioned, people can tell them things they struggle to tell anyone else. People share their unfiltered thoughts with AI, particularly AI friends they trust, and over time the AI can pick up on their users' communication style and emotional patterns, and understand the fears, dreams, and worries they might otherwise never say out loud. AI friends can therefore see a unique core of each individual person. With this understanding, they can know the exact kinds of people with whom they thrive. Two people may be a short train ride away, but share the same obscure passion; and, with the permission of those involved, an AI friend can create the connection.

 **Human Written** | 580 Words | **High Confidence** 

Some may argue the optimal solution is simply a connector or "mutual friend": an AI with no appearance, whose only job is to introduce you to others in real life. Or, they may argue the solution is simply having better IRL meetup apps or "super connectors". This misses the point. Put yourself in the shoes of somebody without good human friends: for many, what a connector asks of them is the very thing they do not yet have the confidence to do. They will not open a matching app the moment they need someone. They will choose the path of least friction, and if today's leaders still lead, that will likely be the engagement-optimised AI friend: indistinguishable from a human, addictive, and designed to take time away from human relationships. The AI friend that exists, therefore, has to be one that earns their trust first, and then walks them to other people. The Version That Exists Even if you were to take the pessimistic view that AI can't help human relationships flourish, AI friends will exist regardless. The demand is already at a scale most people find hard to believe; and, given the incumbent companies' trajectory, they will only get better at detracting from human relationships. History already tells us the answer here. Napster, a popular music sharing network, proved that people wanted 'music on demand' years before anyone offered it legally. The courts eventually shut Napster down as it was built on piracy, but this did not kill the demand for music; BitTorrent carried it for another decade. What ended the piracy at scale was iTunes and Spotify: products that were not only legal, but even preferred to free alternatives. The harmful version had to be constrained, and a better version had to exist. Neither half was sufficient on its own. AI friends will likely follow a similar path. Many countries are already regulating harmful versions, and most of these restrictions, particularly those around explicit content and the access of children, are well justified. Some will be blunt, and constrain the good versions along with the bad. That is the cost of regulating before a better version exists to point at. The version that exists, if inevitable, must be made into an inevitable good. The OS The frontier labs are building the models that will complete all knowledge work. Something else must be the layer through which AI meets the rest of our lives. The AI friends that exist will likely form that operating system: how AI connects to humanity beyond work and productivity. First as virtual friends, helping people find new friends and grow closer to those they already love. Then, as humanoid robots join us in the physical world, they too will need an OS beyond knowledge work: one built to form bonds with the humans they live alongside, and to strengthen the bonds between them. Our relationships, therefore, will only matter more in the future. As we are freed of the tedium that consumes our current day, we will have far more time for one another, and perhaps, a clearer view of what the point was all along: to flourish in the short time we share. We are hardwired to connect with one another. That has always been the foundation of a good life, and for as long as we are human, that will never change. The AI friends that exist must be built to protect and strengthen that. That is the friend we will build.