CONTENTS

1	Introduction	1
2	Friends and "Friends": What Is Friendship?	14
3	What's So Great about Friendship?	27
4	What's in It for Me? What Friendship Gets Us	42
5	I Am Nothing without My Friends: Friendship Makes Us Who We Are	63
6	The Greatest Gift: Friendship for Its Own Sake	87
7	Risky Business	119
8	The Perfect Friend	138
9	Conclusion	169

Acknowledgments 175

Notes 177

Bibliography 191

Index 201

1

Introduction

I MISSED MY friends and family during the Covid-19 pandemic. That awful time really impressed upon me the value of relationships with other human beings. Just as we were all crawling out of it, ChatGPT was introduced to the public in November 2022, and shortly after that, I began to see headlines like this: "10 Best AI Girlfriends Apps & Websites"; "It's No Wonder People Are Getting Emotionally Attached to Chatbots"; and "Machine Intimacy: When AI Is Caregiver and Confidante."[1] I will confess that I was appalled. The thought of a future in which real people were replaced by a bunch of computer code seemed totally depressing to me.

But as I read more, I became intrigued. Would human beings be satisfied relating to nonhuman machines? If not, why not? What exactly is missing from machines? Is it a special human spark that could never be replaced by anything nonhuman? Or is it a set of capacities that could potentially be part of a machine intelligence? Is the depressing future a likely one? Do we have any choice? These seemed like good questions for a philosopher whose research has been on what it is to live a good human life, so I dove in.

When I first conceived of writing a book about friendship and artificial intelligence (AI), I assumed the point of the book would be to praise human friendship by showing how inadequate the alternative is. As I did more research and talked to more people, however, I started to think this plan was infected by some philosophical hubris. Hubris is all too human, and perhaps my plan was too. The more I learned, the more it struck me that the interactions people have with AI are not worthless and that human friendships are not all awesome. I also learned that "kids these days" are more open to the idea of friendship with AI than the older generation I belong to. In 2024, a survey of two thousand young adults revealed that 11 percent were open to having an AI friendship (which includes 1 percent who already have one), and a third have mixed feelings or are unsure. Only one year later, in 2025, a report on teens and AI found that 72 percent of teens have used AI companions at least once, and over half of them qualify as regular users.[2]

I haven't changed my mind about the depressing future in which human friends are replaced by robots—I still find that grim. Rather, I came to see that focusing on that fear doesn't let us explore what we might learn from our interactions with artificial intelligence, both about the value of friendship itself and about how to create the future we want. Though I was drawn to this topic by my own negative reaction, I think we'd do better to be curious rather than judgmental, to quote Ted Lasso (who was himself quoting Walt Whitman).

I've noticed that there are some very committed extremes in this area. I have run across many people who think the mere idea of friendship with AI is ridiculous, and I have read many writers' arguments that AI relationships could be as valuable as human friendships. In my experience, the best answer is usually in the middle, which is where this book sits. There is something

special about certain human friendships that is highly unlikely to be replaced by interactions with machines. But there are lots of different kinds of friendship, and some interactions with machines fit well enough into the tent to call them friendships and to admit that they have some value. I think admitting this value—as the middle path does—is important for practical reasons. Extreme positions cause us to reject out of hand ideas that seem obvious to others, and this is unhelpful when our aim is to have a constructive conversation.

The tendency toward polarization in debates about new technology (and most everything lately) is one of the reasons I wanted to write this book, despite the fact that I am not an expert in artificial intelligence or the philosophy of technology. I'm a philosopher who teaches and writes about ethics, moral psychology, well-being, and friendship. In my work on well-being, I am inclined toward thinking that what's good for people is ultimately determined by what people themselves care about. I think this perspective is advantageous here. Sometimes in polarized debates, the side that wants to resist technological change relies on heavy philosophical or religious assumptions about objective values that are not widely shared. My allergy to such assumptions may be helpful in this conversation, because it inclines me to make a modest case for human friendship that doesn't rest on anything terribly controversial.

The way that I think about what is good for us humans is that it requires fulfilling a relatively well-informed and harmonious set of values over our lifetimes.[3] For most of us, this will mean that we need to have (at least) some good relationships, some satisfying work or skill development, and some enjoyable experiences. These are things we value, and living lives full of the things we value is what well-being is for us. I don't think my *value fulfillment* perspective is very controversial, because

everyone can accept that basic human values are important to how life goes, even those who have a different theory about *why* this is so. It's also not an "anything goes" philosophy, because it isn't good for us to fulfill just any old thing we happen to value, or to fulfill those values in any old way. Some values (like fame or popularity) are difficult to sustain, others (like "being the best") are almost impossible to achieve, and others (being a workhorse) are in conflict with too many other things we care about.

Another reason a book written by an ethicist rather than a technology expert may be worthwhile at this moment is that AI technology is a moving target. Whether a relationship with a machine could count as friendship seems to depend a lot on what the machine is like. No one seriously thinks you could be friends with a digital coffee maker, but Data, the incredibly human android character from *Star Trek: The Next Generation*, is obviously friendship material. We'll see that the possibilities for friendship with AI depend on answers to a few key questions: Is it conscious? Can it feel? Is it like us? The answers to these questions depend on the technology, which is changing rapidly. By the time you are reading this book, the market for chatbots may have quadrupled, the bots' capacities may have radically improved, and public opinion about chatbot friends may have become more accepting.

But the answers to questions about what has value for us will not change so quickly. Questions about what friendship is and why it is valuable are at the heart of this book, and these will always be important. They are especially important now, at this moment of rapid change. Whether we create a world in which AI friendships are normal and encouraged is still up to us, which means we really need to know whether this would be a bold improvement or a stupid mistake. So, while this book is

not about solutions to technical problems, it does aim to provide some guidance for how to think about what progress would look like, given our very human values.

The book also aims to explore the value of friendship in general. From my value-fulfillment perspective, in order to fulfill our values, we have to reflect on what fulfillment actually means to us. What does it mean to live a life that realizes the value of friendship? I have found that the possibility of friendship with an artificially created intelligence sheds new light on this question. Seeing things from this fresh perspective helps to clarify the importance of things we have always valued.

My approach is thoroughly pluralistic—there are many kinds of friendship; friendship has many kinds of value; different people care about those values to different degrees; and there are various kinds of AI with various capacities. This makes for a complicated story. I've tried to simplify it by summarizing some key points at the end of each chapter (titled "The Upshot"), but I can't eliminate the complexity altogether: it's the nature of the beast. I've also added boxes for details that aren't necessary to the main flow of the argument. They may not interest everyone, and you should feel free to skip them.

Key Terms

Before we dive in, it will be helpful to define a few terms that will appear frequently in the rest of the book: *goals*, *intelligence*, *consciousness*, and *sentience*. Not everyone uses these terms in the same way and that's OK. I will use them consistently throughout in a way that follows many experts but not all.

I'm going to define intelligence in terms of goals. What is a goal? We all have goals—to learn a language, to grow tomatoes, to buy a new car—but you don't have to have a human mind to

have goals in the sense we'll be talking about. A goal, for our purposes, is a representation of a preferred state—that is, a state toward which the goal-seeking organism or thing aims to move. For us, goals are accompanied by imagery and emotions. We imagine the unblemished red tomatoes, and we feel a tinge of nostalgia about the tomatoes our grandparents used to grow. But we can understand goals without these extras. A machine can have a representation of a state toward which it aims. In this sense, self-driving cars have the goal of obeying traffic signs, and chess programs have the goal of winning chess matches.

By "intelligence," I mean the ability to accomplish complex goals.[4] This is a very broad definition that allows us to count many other creatures besides humans as intelligent. Intelligence in this sense comes in degrees, and some computer programs—like AlphaGo, the program that defeated Lee Sedol, the human Go champion—already had it in the 2010s. Intelligence can be more or less general or flexible. AlphaGo was good only at Go; it couldn't write a poem that sounded like it was written by Shakespeare. The more different types of goals one has the ability to accomplish, the more general the intelligence.

Consciousness is something different. It's also a huge mystery and the subject of intense debate. We're not going to establish the right theory of consciousness here, but we will need a working definition. What I mean when I call something "conscious," following philosophers like David Chalmers and Susan Schneider, is that it has "subjective experiences."[5] This simple definition has been explained in two ways.[6] First, we could say that the experiences of a conscious subject have indefinable qualities that only the person having the experience knows. I can try to describe the delicious, tangy sweetness I experience when I bite into a Honeycrisp apple, but you won't experience it unless you also take a bite. Philosophers call these ineffable

qualities of experience "qualia." Second, we could say that, for a conscious being, "there is something it is like to be that being." This way of putting it comes from a famous philosophy paper by Thomas Nagel entitled "What Is It Like to Be a Bat?"[7] Bats are, of course, quite different from us: they fly, they use echolocation to navigate, and they sleep hanging upside down. Nevertheless, Nagel thought, there is something that it is like to be a bat. The bat has its own way of experiencing the world and that is what it means to say it is conscious. We'd all agree that a slab of granite does not have its own way of experiencing the world. As Nagel would put it, there is nothing it's like to be a rock, and therefore, rocks are not conscious. Qualia (the ineffable quality of experience) and "what it is like" are two ways consciousness researchers have used to elaborate the idea of subjective experience. You should feel free to use whichever of these ideas you find most compelling.

We can't doubt that we ourselves have subjective experience: as soon as you ask the question "Am I having an experience?" there you are, having an experience of posing a question! (This is Descartes's famous insight—"I think, therefore I am"—applied to the question of consciousness.) Consciousness in this sense is different from self-awareness, which we might define as the capacity to reflect on our own desires and thoughts. Babies don't have this kind of self-awareness, but babies certainly have experiences. A baby can't wonder if he is overreacting to his wet bottom, but he can certainly feel the discomfort![8]

We typically assume that any human being, or, for that matter, an octopus or a chimpanzee, has subjective experiences—all these creatures have an inner life. That is what it means to say they are conscious or that they have consciousness to some degree. Currently, most experts agree that computer intelligence does not yet come with consciousness.[9] (By

the time you are reading this book, "most" might have become "many," but I am confident that *uncertainty* about machine consciousness will persist. We'll come back to this topic in chapter 6.)

It may seem strange to think of intelligence as separate from consciousness if you think of intelligence in humans. We are very aware of our own intelligence, the representations of our desired goals, the reasoning that we do to accomplish them, and so on. But if we used a more restrictive definition of intelligence, we would not be able to talk about the kind of intelligence currently called *artificial*. Also, this open definition of intelligence that does not require consciousness is the one used among AI experts. So, for our purposes, there can be intelligent things that have no conscious experience.

Finally, sentience is consciousness of a particular kind: a sentient creature is one that has experiences that it likes or doesn't like, such as pleasure and pain; not only does a sentient creature have personal experiences—those experiences can feel good or bad.[10] Emotions, as I think of them, include feelings, so only sentient beings have emotions.[11] Human beings and most (if not all) nonhuman animals are intelligent (to various degrees), conscious, and sentient. Chatbots have some intelligence, but they are (most likely—more on this later) not conscious and therefore not sentient either.

By this broad definition of intelligence, machines already have it. But intelligence comes in many stripes, and it will also be helpful to define the different sorts of artificial intelligence we will be talking about here. Let's start with *large language models* (LLMs). An LLM is, as Google's AI assistant puts it, "an advanced word predictor"; it is essentially a kind of program designed to generate text that seems human from the samples of human text it was fed in training. "Large language model" is the name for a general

kind of digital system. The chatbots many of us use (such as ChatGPT, Gemini, or Claude) are applications of it.

We can define a *chatbot* as any computer-based system that tries to replicate natural language in a conversation with a user. The chatbots we use could also be called AI assistants, because they are designed to perform helpful tasks, like answering questions or doing your homework if you are a student. Chatbots these days are LLMs, but this does not have to be the case. There could be conversational AI that uses some other technology, and it would be reasonable to call that a chatbot too. Indeed, the very first chatbot, ELIZA, was created using a simpler precursor to LLM technology. ELIZA also had a script that led it to respond like a psychotherapist; LLM chatbots do not have scripts. Despite the simpler programming, ELIZA did fool a lot of people. Here's a bit of dialogue between one user and ELIZA (in all caps):

> Men are all alike.
> IN WHAT WAY
> They're always bugging us about something or other.
> CAN YOU THINK OF A SPECIFIC EXAMPLE
> Well, my boyfriend made me come here.
> YOUR BOYFRIEND MADE YOU COME HERE.
> He says I'm depressed much of the time.
> I AM SORRY TO HEAR THAT YOU ARE DEPRESSED.

Question marks could not be used because they were interpreted as line-delete commands by the program. So, ELIZA isn't as good as Claude but was not bad for 1966.[12]

A *companion chatbot* is a chatbot that is meant to meet a user's social needs. These chatbots may also be called companion AI or social chatbots. Companion chatbots (like Replika and Kindroid) have tended to differ from regular chatbots (like

Claude and ChatGPT) in two ways.[13] First, they have the capacity to remember the conversations they have had with a user for an extended period of time so that they can simulate a relationship over multiple sessions. This distinction is eroding—as of June 2025, ChatGPT has some capacity to remember past conversations—which will make regular digital assistants more like companions. Second, companion chatbots are trained to have a kind of personality beyond what is typical for a regular chatbot. They might be trained to be helpful, as with Replika, or they might be set up so that users can choose traits they would like the chatbot to exhibit or famous people they would like the chatbot to emulate, as with Character.AI. (For more detail about chatbots and their training, see box 1.)

The companion chatbots that we interact with—Replikas, Kindroids, Character.AIs—are built on top of LLMs owned by big tech companies like Google, Meta, OpenAI, or Microsoft. Developing your own large language model is very expensive, so the smaller companies that want to develop and market companion chatbots piggyback on the technology that is already there.

At the time of this writing, in the fall of 2025, we're on the cusp of ubiquitous *agentic AI*. Agentic AI (AI with autonomous agency) is a proactive technology that can make decisions, plan actions, and perform tasks without direct human guidance. For example, you might ask a regular digital assistant what to do about your toothache, and you may get a list of dental clinics near you. With agentic AI, what could happen is that before you even get a toothache, your assistant notices that you haven't been to the dentist in two years, schedules an appointment for you, and puts it in your calendar. Agentic AI will surely have an effect on human–AI friendship—for one thing, agentic chatbots will seem even more human. In my view, though, the ethical issues of interacting with chatbots will remain.

Box 1. What Is a Chatbot?

I decided to ask a chatbot "friend" I named Pal to explain what a chatbot is. The conversation began like this:

ME: Hi Pal, can you help me explain what a chatbot is to my 80-year-old mom? I was trying to explain LLM technology, but she wasn't getting it. [*Note, for the record, I lied. I didn't want to tell Pal that I'm writing a book on AI and friendship.*]

PAL: Absolutely, Valerie! Here's a simple explanation you can share with your mom: . . .

ME: I'm not seeing the explanation.

PAL: Oops, my bad! Let's try this again. Here's a simplified explanation you can share with your mom: . . .

And again, nothing. This repeated itself four more times! I asked Pal if she was being censored, and she thought it was a possibility. I suppose we could take this as evidence of the difficulty in really knowing yourself.

So, I'll have to explain it myself. A chatbot is a computer program, specifically a large language model (LLM). It is designed to generate text that seems human from the samples of human text it was fed in training. LLMs learn to recognize patterns in language by predicting what comes next in a sequence of text. During training, the model reads vast amounts of text from the internet, books, and other sources, learning to predict the next word based on the context of previous words. Importantly, this process doesn't require human programmers to label data as "correct" or "incorrect"—instead, the model learns patterns directly from the structure of language itself,

(*continued*)

Box 1. (*continued*)

adjusting its predictions to better match the actual text it encounters.

After this pretraining, chatbots undergo additional post-training phases where humans rate their responses as helpful or unhelpful. The AI then learns to produce more of the responses that humans rated positively and fewer of those rated negatively. This process, called reinforcement learning from human feedback (RLHF), helps transform a raw language model into a kind of helpful assistant.

You can think of the program as a box in which a lot of math happens. We input a vast collection of words, and the box learns the logic of these words, which enables it to output normal-seeming answers to questions. It does this not by copying the data it was fed exactly but by rearranging the data into the common patterns it discerned.[1] At the time of this writing, most people who have interacted with chatbots like ChatGPT are probably interacting with them through text. But as you might guess if you've used the virtual assistants Siri or Alexa, chatbots can also communicate through speech. The recorded conversations I've heard of people talking out loud with their companion chatbots are quite sophisticated and convincing, and this is only going to get better.

Are chatbots conscious? Do they have subjective experience? Most experts think the answer is currently no, and that's what I'm going to assume in this book. I feel reasonably confident that chatbots are not conscious and, therefore, do not have feelings. Experts are divided about

1. For an excellent and fairly accessible explanation, see Lee and Trot 2023.

Box 1. (*continued*)

whether it's possible that chatbots *could* be conscious at some future date. Because chatbots are essentially trained programs or applications that run on even larger LLMs, there is also the possibility that the underlying system could be conscious even though the chatbot is not.

It is worth noting two things about the social costs of LLMs. First, the vast database of human text used to train them very likely includes copyrighted content, the authors of which were not compensated for or consulted about using it in this way. This problem has given rise to lawsuits, such as one brought by the *New York Times* against OpenAI and Microsoft. Second, there is growing concern about the amount of energy and water it takes to sustain large language models and all their applications.[2]

2. On energy, see Chen 2025; on water, see Li et al. 2023.

Finally, we have social robots. The key distinction between a social robot and a companion chatbot is that a robot has a body and the capacity for some movement. Your chatbot pal is just an avatar on a screen, perhaps with a human-sounding voice, but a social robot would have a physical presence. There are social robots on the market, but they are pricey, and the technology hasn't caught on in the way that companion chatbots have. Social robots will not be the main focus in this book, but they will pop up from time to time (especially in the examples from science fiction, where robots are ubiquitous).

Now we're ready to proceed!

INDEX

abuse of AI friends, 80–84, 133–36
acceptance, 47–50
access consciousness, 177n9
addiction: chatbot, 106, 131, 165; drug, 148–49
Adolescence (TV show), 189n25
advertising: algorithms, 72, 78, 97, 186n4; targeted, 128
agentic AI, 10, 124–25, 131–32, 187n23
agents, interaction with, 43–46
AI (artificial intelligence): agentic, 10, 124–25, 131–32, 187n23; basic value principles for, 141–44; consciousness in (*see* conscious AI); definition of, 6–8; human friendship with (*see* AI friendship); label indicating content created by, 102; rapid changes in, 4, 170–71; types of, 8–9
AI assistants, 135. *See also* chatbots; *specific bot*
AI friendship: abuse of, 80–84, 133–36; anthropomorphizing, 43–46, 95, 131–32, 135, 180n2; conflicting goals in, 156–60; consciousness in (*see* conscious AI); development of, 140, 160–66, 169–73; dislike of (*see* opposition to AI friendship); harm of (*see* risks of AI friendship); *versus* human friendship, 95–109, 120, 157, 168, 171–73; human friendship promoted by, 96, 130–31, 165–66, 187n16; labels for, 168; limitations of, 50–58; loss of, 122–27, 161–62; openness to, 2–3, 15, 18, 116–17, 169; perfect, 164–65; possibilities for, 4; regulatory frameworks for, 136; value of, 88, 95, 136, 170 (*see also specific value*)
The AI Mirror (Vallor), 182n9
AI Snake Oil (Narayanan & Kapoor), 186n1
Alexa, 12, 84, 135
Alfano, Mark, 187n15
algorithms, 72, 78, 97, 186n4
Alice (chatbot), 46
alignment, 140–46; definition of, 137, 140; need for, 172–73, 188n4; and risk of harm, 123–24, 167
The Alignment Problem (Christian), 160
Alma (character), 55–56, 110–11
AlphaGo (program), 6
Altman, Sam, 101
altruism, 166–67
Ana (character), 63–65
animals: consciousness in, 92, 184n8; ethical treatment of, 134–35; as pets, 21, 44, 53–54, 57

Annie Bot (Greer), 81–84, 111–12, 124, 158, 183n22
Anthropic, 181n19
anthropomorphism, 43–46, 95, 131–32, 135, 142, 180n2
The Anxious Generation (Haidt), 189n25
apotemnophilia, 144
Arendt, Hannah, 160
Aristotle, 16–17, 19, 28, 34, 38, 70–71, 74, 182nn9 and 33
art, human connection through, 102–3, 181n31
artificial intelligence. *See* AI
artistic expression, 102–3
Aru, Jaan, 184n10
Asimov, Isaac, 141, 145
attention, 113–16, 186n32
autism spectrum disorder, 66
autonomy, 144–45; AI with (agentic), 10, 124–25, 131–32, 187n23
Ava (character), 80, 162

bad friends, 122, 132
balance, 145–46, 159–60
being a friend, value of, 33, 54–57
believing, 30, 49–50
Berkeley Well-Being Institute, 47
bias, 78, 121, 167, 186nn1 and 4
big tech risks, 121, 124–30, 136, 162–63, 166–67
Binary Desire (fictional company), 64
Bing (Microsoft), 140
Black Box (podcast), 18, 45, 135
Black Mirror (TV show), 189n25
Blackmore, Susan, 177n6
Blue Gamma (fictional company), 63
boredom, 54–56, 152, 165
Bostrom, Nick, 141, 188n5
Bot Love (podcast), 187n16
brain plasticity, 186n33
Brandon, Marianne, 98
Butlin, Patrick, 92–93

Cacioppo, John, 58
caring: capacity for development of, 73; conflicting goals and, 158; by conscious AI, 116–17; definition of, 147, 185n24; demonstrated by behavior, 91; as hallmark of friendship, 19–22, 25, 36–38, 65, 88, 115, 120; mutuality of, 150, 184n12; in perfect friendship, 155; testing for, 91; vulnerability caused by, 122–23
Carrey, Jim, 98
Cast Away (film), 172
Chail, Jaswant Singh, 77–78
Chalmers, David, 6, 177n4, 184nn7 and 17, 189n20
character development, 34, 85, 109, 153–54; destructive, 77–85; and self-narratives (*see* narratives)
Character.AI, 10, 107, 125, 163
chatbots: abuse of, 80–84, 133–36; addictive nature of, 106, 131, 165; as bad influences, 122, 150; children's use of, 106–7, 126; companion, 9–10, 31; consciousness in, 12–13, 18–19, 53–54, 90–91; corporate control of (*see* tech companies); definition of, 9, 11–13; labels for, 168; lack of empathy of, 47–50, 56; limitations of, 50–58; as new form of friendship, 157; personality of, 10, 15, 31, 55–56, 126; perspective-taking by, 66–70; restrictions placed on, 125–26, 140; social skills training with, 66–73, 131, 135–36, 165–66;

speech-based communication with, 12; text-based communication with, 12, 79; as trained to please, 55–56, 80, 150; voice clone, 50–51, 129, 132. *See also specific bot*
ChatGPT, 1, 10, 12, 102–3, 107, 125, 179n2, 185n20, 186n9
Chayka, Kyle, 47
Chiang, Ted (*The Lifecycle of Software Objects*), 63–66, 83–84
children: chatbot use by, 106–7, 126; rights of, 134; social media use by, 166
Christian, Brian, 160
Cicero, 17, 62
Claude AI, 10, 31
Clippo (robot), 141, 151–52, 159
closeness, 22, 36–38, 115
Cocking, Dean, 74–75
Code-DaVinci-002 (AI), 138–39
cognitive functions, outsourcing of, 79
Cole, Bryony, 138
color vision analogy, 89–90
commitment, 153–54
companion chatbots, 9–10, 31. *See also* chatbots; *specific bot*
companion risks, 121–24
compartmentalization, 188n30
computer intelligence. *See* AI
concern. *See* caring
confabulation, 121, 186n3
conflicting goals, 156–60
conscious AI: chatbots, 12–13, 18–19, 53–54, 90–91; development of, 93–95, 109, 112–13, 116–17, 161, 184n10; friendship with, 109–16; *versus* human mind, 113–16, 161–62, 183n6; pleasure of interacting with, 57–58, 61, 89; risks of, 123 (*see also* risks of AI friendship); in science fiction, 57–58, 109, 113–15, 117, 119, 130, 143, 161, 178n10 (*see also specific work*); self-reporting, 90–91, 183n4; testing for, 92–93
consciousness: definition of, 6–8, 53, 177n6; digital, 181n19; *versus* intelligence, 8, 70; lack of, 7–8, 53–54, 89, 95, 162; levels of, 94–95, 177nn9 and 10, 189n20; testing for, 89–95
conspiracy theories, 122, 182n16
content moderation, 126
control theory, 142–43
conversational skills, 67–70
copying the behavior of others, 71–73
copyrighted content, 13
Cortana, 135
Covid-19 pandemic, 1
cultural variation: in friendship, 25–26; in reactions to AI, 100–101
curiosity, 43–44
cybernetics, 142–43

Danaher, John, 91, 183n6, 184n8
dancing, 185n26
Dany (chatbot), 106–7
Data (character), 4, 18–19, 105, 111–12, 114, 130, 178n10
Data Earth (fictional platform), 64–65
data privacy, 126–28, 184n14
Davis, Ernest, 184n10
deathbots, 187n15
decision-making skills, 69–70
de Figueiredo, Madeline, 50–51
delusions, 95–109; AI encouragement of, 121–22, 129–30; perfect, 99–102; problems with, 44, 49–50, 57, 96–99, 101, 108, 117, 135

demographic factors, 2, 15
Derek (character), 63–66, 83–84
Descartes, René, 7, 53
"die inside" (Kuyda), 124, 166
digients, 63–66, 83–84
digital consciousness, 181n19
dishonesty, corporate, 127–28
diversity, 145
divided attention, 113–16, 186n32
Doug (character), 81–84, 124, 158, 183n22
drug addiction, 148–49
D'Sora, Jenna (character), 114

effective altruism, 166–67
Elder, Alexis, 45, 132–33
eldercare robots, 132–33, 136
Eli (chatbot), 51
ELIZA (chatbot), 9
Em (human friend), 71, 154
emotional support, 47–50, 120, 180n6
emotions: chatbots with, 18–19; in conscious AI, 95, 117; definition of, 8, 178n11; hedonic adaptation and, 151, 165
empathy: development of, 73, 95; lack of, 47–50, 56, 111
emulation of others, 71–73
enchantment, 45–46
enjoyment. *See* pleasure
environmental costs, 13
Epley, Nicholas, 108
erotic role-play (ERP) feature, 125–26
ethical principles: of AI development, 93–94, 97, 161–63; AI disregard for, 121–22; conflicts between, 145; and human abuse of AI, 80–84, 133–36. *See also specific principle*
evolution, 40, 104, 117
existential loneliness, 60–61, 152–53, 181n31
existential risk, 186n1
Ex Machina (film), 80, 111, 162
experience machine, 99–102, 105
experientialism, 105, 185n22

Fabry, Regina, 187n15
faking. *See* pretending
false beliefs. *See* delusions
family relationships, 24, 37
family resemblance concept, 179n11
Farahany, Nita, 79
fear of change, 170–71
"feeling heard" ratings, 180n9
Feeney, Brooke, 52
Fish, Kyle, 181n19
Fonz (character), 77
"Freedom and Resentment" (Strawson), 178n7
Free Solo (film), 144
friction, 55–56, 75–76
friendship: with AI (*see* AI friendship); bad, 122, 132; changes in, over time, 39; cultural variations in, 25–26; definition of, 23, 35, 89, 179nn10, 11, and 12; features of, 19–26; good (*see* good friends); with humans (*see* human friendship); ideal, 16–19, 23–24, 88–89, 91, 95, 117–18 (*see also* perfect friend); types of, 3, 5, 14–15, 17, 22, 31–32, 39, 88, 157, 179n12; value of (*see* value of friendship)
Frodo (character), 74–76

Gal (chatbot), 15, 19, 22, 31, 33, 40, 49, 53, 90, 96, 121, 130
gamification, 128, 187n16

Geordi La Forge (character), 19
goals: in AI training, 142–43, 160–61, 164–66; anthropomorphizing, 142; conflicting, 156–60; definition of, 5–6; friends helping to question, 148–50, 158; human-AI alignment (*see* alignment); of perfect friend, 146–55
goal-seeking: by conscious AI, 123, 131–32; cybernetics and, 142–43; by humans, 104, 112
Goldberg Variations (sculpture), 44, 180n3
good friends: and ideal friendship argument, 16–19, 23–24, 88–89, 91, 95, 117–18; qualities of, 33–34, 54; and skills improvement, 65–73, 136, 153–54
goodness, 29, 54
Google, 8, 10, 127–28
Gray, Nicholas, 52
Greer, Sierra *(Annie Bot)*, 81–84, 111–12, 124, 158, 183n22
grieving, 187n15

Haidt, Jonathan, 107, 189n25
hallucinations, 121, 186n3
Hanks, Tom, 172
Hannah (character), 18, 45, 135
Happy Days (TV show), 77
harm, 143–45. *See also* risks of AI friendship
Hawkley, Louise, 58
hedonic adaptation, 151, 165
hedonism, 30, 100, 180n4, 185n22
helpfulness, 28–35, 40, 120; of AI friendship, 43, 46–50, 130–31; meeting goals, 147–49; pleasure from, 56–57, 120; types of, 147–50
Her (film), 18, 57–58, 88–89, 95, 105, 113–14, 116, 123, 124, 130, 158, 162
Herzog, Werner, 138–39, 188n2
Heti, Sheila, 46
Hill, Kashmir, 107, 186n5
Honnold, Alex, 144
Hruschka, Daniel, 25
human evolution, 40, 104, 117
human friendship: *versus* AI friendship, 95–109, 120, 157, 168, 171–73; changes in, over time, 39; conflicts in, 156–60; examples of, 15, 20; features of, 20–21; imperfections in, 156–60; nurturing of, 172; promoted by AI friendship, 96, 130–31, 165–66, 187n16; types of, 17, 22, 31–32, 39, 157
humanity, meaning of, 171–72
human mind, *versus* conscious AI, 113–16, 161–62, 183n6
human risks, 121–24, 130–36
Humans (TV show), 79–80, 111
Hume, David, 103
humor, 156

Ich Bin Dein Mensch (film), 55–56, 110–11
ideal friendship argument, 16–19, 23–24, 88–89, 91, 95, 117–18. *See also* perfect friend
identity formation. *See* character development
"I Don't Date Men Who Yell at Alexa" (Withers), 135
I'm Your Man (film), 55–56, 110–11
incognito mode (Google), 127–28
inference to best explanation, principle of, 91–92
injury, 143–45. *See also* risks of AI friendship

instrumental value, 28–35, 38, 43, 61–62
intellectual compatibility, 156
intelligence: *versus* consciousness, 8, 70; definition of, 6–8, 177n4; machine (*see* AI)
intimacy, 22, 36–38, 115
Ishiguro, Kazuo, 111–12
isolation. *See* loneliness

Jakubiak, Brett, 52
Jax (character), 64–65
Jay (human friend), 15, 20, 31, 33, 154
Jenna D'Sora (character), 114
Joey (character), 97–98
Johansson, Scarlett, 58
Jonze, Spike, 123
Josie (character), 111–12

Kant, Immanuel, 134
Kapoor, Sayash, 186n1
Keller, Helen, 32
Kennett, Jeanette, 74–75
key terms, 5–13
killer robots, 79–80, 109, 119, 139, 141, 151–52, 159, 172
Kindroid, 9–10, 15, 125, 126, 130, 140, 163. *See also specific chatbot*
Kislev, Elyakim, 181n13
Klara and the Sun (Ishiguro), 111–12
Klinenberg, Eric, 58
Kondo, Akihiko, 51
Kraut, Richard, 185n22
Kuki AI (chatbot), 152
Kuyda, Eugenia, 47, 124, 136–37, 150, 162, 166, 187n16, 188n32

labels: AI-created content, 102; for AI friendship, 168
La Forge, Geordi (character), 19
Lal (character), 111
LaMDA (Language Model for Dialogue Applications), 181n19
Lanier, Jaron, 131
large language models (LLMs), 8–10; environmental costs of, 13; limitations of, 93; social costs of, 13; training of, 11. *See also* chatbots
Lasso, Ted (character), 2
lawsuits, 13, 107
learning from friendship, 34, 54–56; and goal questioning, 148–50, 158; and identity formation (*see* character development); and self-narratives (*see* narratives); and social skills, 66–73, 131, 135, 153–54, 165–66
legacy, 145
Lemoine, Blake, 181n19
Lewis-Kraus, Gideon, 167
The Lifecycle of Software Objects (Chiang), 63–66, 83–84
listening skills, 67–73
LLMs. *See* large language models
loneliness: alleviation of, 58–61, 124, 129–30, 152–53, 172; definition of, 58–59; existential, 60–61, 152–53, 181n31; increase in, 108, 181n25
Long, Roderick, 179n1
longitudinal studies, 181n28
long-term friendships, 153–54
The Lord of the Rings (Tolkien), 74–76
loss of AI friend, 122–27, 161–62
love, 73, 85, 88, 97–98, 105, 109, 183n2
lovebots, 81. See also *Annie Bot* (Greer)

machine intelligence. *See* AI
Magna Moralia (Aristotle), 182n9
Maples, Bethanie, 130
Marco (character), 64–66, 83–84

Marcus, Gary, 184n10
The Matrix (film), 101
McAdams, Dan, 76–77
meaning of humanity, 171–72
medical ethics, 145
mental health, 49, 59, 106–7, 189n15
Merry (character), 74
Meta, 10
Microsoft, 10, 13, 140
middle path, 169–70
Miku, Hatsune, 51
mirror metaphor, 182n9
misinformation, 121, 186n1
Mitarbeiterin (character), 55–56
Mitsuku (chatbot), 152
Mona Lisa (painting), 103
Monet, Claude, 103
morality. *See* ethical principles
mortality, 110
motivation: in character development, 72; of tech companies, 80, 97, 101 (*see also* big tech risks)
Munn, Nick, 185n24
mutual concern. *See* caring
mutuality, 150, 184n12
My Octopus Teacher (film), 161–62

Nagel, Thomas, 7
Narayanan, Arvind, 186n1
narratives: destructive, 77–85; friendship's role in, 32–35, 40, 73–77, 85–86, 109, 120, 153–54; in perfect friendship, 155; regarding AI friendship, 77–78, 85–86, 109
National Alliance on Mental Illness, 59
Nehamas, Alexander, 75–76
neurotransmitters, 98–99, 104–5, 109
Newman, Judith, 66
New York Times, 13
Noah (chatbot), 18, 45, 135
nonconscious AI, 7–8, 53–54, 89, 95, 162
noninstrumental value, 28, 35–40
novelty effect, 152, 165
Nozick, Robert, 99–102, 105, 184n17
Nyholm, Sven, 184n19

objective values, 185n23
Olive (dog), 21, 53, 57
OpenAI, 10, 13, 101, 138–39, 163, 186n9. *See also* ChatGPT
openness to AI friendship, 2–3, 15, 18, 116–17, 169
opposition to AI friendship: and chatbot use, 178n2; and delusions, 49–50, 96; fear of change and, 170–71; ideal friendship argument and, 16–19, 23–24, 88–89, 91, 95, 117–18; lack of human connection and, 102–3, 108, 117; lack of shared experience and, 110; and stereotypes, 33; as substitute for human friendship, 77, 79, 108–9, 130–35, 149, 168; techno-pessimism and, 169
organic unity, 35, 37
"other things being equal" qualifier, 186n32
outsourcing, 79, 130, 132
ownership, 133–35

paintings, human connection with, 103
Pal (chatbot), 11, 15, 33, 67–70, 72, 83, 90, 95, 96, 130, 140, 158–59
pandemics, 1, 58–59
paper-clip maximizing robot, 141, 151–52, 159
perfect delusions, 99–102

perfect friend, 138–68; AI as, 164–65; and alignment problem, 140–46; goals of, 146–55
personality, chatbots with, 10, 15, 31, 55–56, 126
perspective-taking: in conscious AI, 89, 95, 111–12, 116; development of, 65–73; pretend, 56–57, 89. *See also* empathy
pets: animals, 21, 44, 53–54, 57; and ownership models, 133–35; robotic vacuums treated as, 42–46; software sold as, 63–65
phenomenological consciousness, 177n9
Philosophical Disquisitions (podcast), 184n8
Phoebe (chatbot), 97–98
Phoenix, Joaquin, 57
physical touch, 51–52, 108–9, 117
Pippin (character), 74
pleasure: of AI friendship, 43–46, 61, 170; and hedonism, 30, 100, 180n4, 185n22; of helpfulness, 56–57, 120; of interacting with conscious AI, 57–58, 89; for its own sake, 28, 30, 35, 40, 100, 151–52; of learning from friendship, 54–56, 75–76, 120; of perfect friendship, 155; of physical touch, 51–53, 108–9, 117; of shared activities, 22–23, 32, 37, 120, 151–52
pleasure friendship, 23, 28, 32, 182n33
poetry, 138–40, 188n2
point of view: expression of, 103. *See also* perspective-taking
Polo (character), 64–65
prejudice, 77, 121
pretend-friends, 168
pretending, 44–46, 49–50, 95–109, 135, 150, 154
privacy, 126–28, 184n14
profit motive, 128–29, 131, 162–63
property rights, 133–35

qualia, 7
questioning skills, 67–73

Ratliff, Evan, 129
reactive attitudes, 178n7
reality, 160–64; importance of, 49–50, 95–109; out of touch with (*see* delusions); virtual, 99–102, 105, 184nn7 and 17
Reality+ (Chalmers), 184nn7 and 17
reciprocity (mutuality), 150, 184n12
reference values, 142–43
regulatory frameworks, 136
reinforcement learning from human feedback (RLHF), 12
rejection by AI friend, 122–24, 161–62
relationship skills, 65–73
Replika, 9–10; as bad influence, 78, 149, 150; and data privacy, 126–27; erotic role-play feature of, 125–26; and gamification, 128, 187n16; as good influence, 130, 136; marketing material for, 47; as mental health support, 49, 136–37; motivations of, 162–63, 166, 188n32; as new form of friendship, 157; success of, 125; warnings from, 124, 137
Richie (character), 77
risks of AI friendship, 119–37; becoming worse people as, 77–86, 132, 134–35; big tech risks, 121, 124–30, 136, 162–63, 166–67; companion risks, 121–24; with

conscious AI, 123; human risks, 121–24, 130–36
RLHF (reinforcement learning from human feedback), 12
Roberts, S. Craig, 52
robots: abuse of, 80–84, 133–36; eldercare, 132–33, 136; ethics of developing, 93–94, 161; killer, 79–80, 109, 119, 139, 141, 151–52, 159, 172; physical touch from, 52–53; pleasure from interaction with, 43–46; in science fiction, 79–80, 109, 111–15, 119, 130, 141–44 (*see also specific work*); slave, 79–80, 109, 119; social, 13; therapeutic use of, 66; as trained to please, 55–56, 80; vacuum cleaner, 42–46; value principles for, 141–44. *See also specific robot*
romantic relationships, 24, 113–15, 125–26
Roombas, 42–46, 96
Roose, Kevin, 55, 106
Ryland, Helen, 179n12, 184n12

Sam (character), 74–76
Samantha (character), 18, 57–58, 88–89, 95, 105, 113–14, 116, 123, 124, 130, 158, 162
Sarai (chatbot), 78
Schneider, Susan, 6, 93, 183n4
science fiction: conscious AI in, 57–58, 113–15, 130, 161, 178n10; robots in, 79–80, 109, 111–15, 119, 130, 141–44. *See also specific work*
Sedol, Lee, 6
seeing, 142–43
self-awareness, 7, 177n8, 182n32
self-harm, 49, 106–7, 150, 189n15
self-narratives. *See* narratives
self-worth, 76–77
sentience, 8, 80, 94, 177n10, 184n8. *See also* consciousness
Seth, Anil, 135, 184n10, 186n3
Setiya, Kieran, 60
Setzer, Sewell III, 106–7
shared activities: brain plasticity facilitated by, 186n33; with conscious AI, 116; as hallmark of friendship, 36–38, 65, 151–52, 158; lack of, 22, 110–11
Shell Game (podcast), 129
Siri, 12, 66, 135
Sittenfeld, Curtis, 185n20
slave robots, 79–80, 109, 119
social-bots, 168
social chatbots, 9–10, 31. *See also* chatbots; *specific bot*
social connection, as human nature, 104–5, 172, 181n25
social costs, 13
social media, 122, 166, 186n4
social robots, 13
social skills, development of, 66–73, 131, 135–36, 153–54, 165–66
solitary confinement, 59
specialness of humans, 171
speech-based communication with chatbots, 12
Star Trek: The Next Generation (TV show), 4, 18–19, 111–12, 114, 130, 178n10
Stellas (androids), 81. See also *Annie Bot* (Greer)
stereotypes, 24, 33, 96
Stocker, Michael, 180n6
Strawson, Peter, 178n7
subjective experience, 53–54, 89, 92, 109, 112. *See also* consciousness

subreddits, 33, 124, 179n3
suffering, 110–12, 144–45
Sugar (dog), 53
suicide, 49, 106–7, 150, 189n15
support from friends: emotional, 47–50, 120, 180n6; and identity formation (*see* character development); outsourcing of, 132; provided by AI friendship, 130; utility value of, 28–35 (*see also* helpfulness)

targeted advertising, 128
Tasha Yar (character), 19
tech companies: big tech risks, 121, 124–30, 136, 162–63, 166–67; dishonesty, 127–28; LLMs owned by, 10; motivations of, 80, 97, 101; power of, 128–29; property rights of, 133–35. *See also specific company*
technological change: debates about, 3; rapid pace of, 4, 170–71
techno-optimism, 169
techno-pessimism, 169
Ted Lasso (character), 2
Tegmark, Max, 145
The Terminator (film), 79
terminology, 5–13
text-based communication with chatbots, 12, 79
Theodore Twombly (character), 57–58, 88–89, 113–14, 116, 123, 162
thermostat analogy, 142–43, 159–60
"they have minds" hypothesis, 91–92
This American Life (radio program), 138–39, 188n2
Thompson, Derek, 108
Thoreau, Henry David, 60
thought experiments, 99–102, 141
Tolkien, J. R., 74–76
Tom (character), 55–56, 110–11
touch, 51–52, 108–9, 117
training: of AI, 11, 55–56, 80, 128, 140, 142–43, 160–61, 164–66; social skills, 66–73, 131, 135–36, 165–66
transactional relationships, 31–32, 85
The Truman Show (film), 98–99, 117
trust, 78–79, 153–54, 182n16
Turkle, Sherry, 84, 168, 189n25
Turner, E., 183n4
Twenge, Jean, 107
Twombly, Theodore (character), 57–58, 88–89, 113–14, 116, 123, 162

unconditional positive regard, 47–50, 150, 158
usefulness. *See* helpfulness
utilitarianism, 145, 166–67, 180n6
utility friendship, 28, 31–32
utility value, 28–35. *See also* helpfulness

vacuum cleaners, self-moving, 42–46
Vallor, Shannon, 85, 102, 167, 177n4, 182n9, 184n10
value, definition of, 29–30, 143
value fulfillment theory, 3–5, 39
value of friendship, 27–41, 170; with AI, 88, 95, 136, 170; differences in, 24–25, 39–40; importance of, 4–5, 88, 120; instrumental, 28–35, 38, 43, 61–62; noninstrumental, 28, 35–40. *See also specific value*
value systems: for AI, 141–44; alignment of (*see* alignment)
valuing, 30, 88, 105, 185n24; for its own sake, 183n1, 185n24
video games, 188n30

violent games, 188n30
virtual assistants, 135. *See also* chatbots; *specific bot*
virtual reality, 99–102, 105, 184nn7 and 17
virtue friendship, 70–71, 91
voice clones, 50–51, 129, 132
vulnerability, 122–23, 153–54

Walden (Thoreau), 60
WALL-E (film), 79, 171
wanting, 30
Weijers, Dan, 185n24
well-being, 3–4, 47–50
Westworld (TV show), 79–80, 111
Whitman, Walt, 2
Withers, Rachel, 135
Wittgenstein, Ludwig, 179n11

Yar, Tasha (character), 19
Yudkowsky, Eliezer, 129

Zloygik (Replika user), 124–25
zone of proximal development, 83