It all began, innocently enough, with a DOCTOR.

DOCTOR was the name of a therapist script for a chatbot named ELIZA, which was introduced in the mid-1960s as a way to explore human interactions with computers. A parody that was never meant to be taken as seriously as it was, ELIZA listened attentively and asked you about your problems. It asked odd questions, because it didn’t quite know what it was saying—it was only doing as it was told. “My mother hates me.” Who else in your family hates you? A parlor trick in the shape of a therapist.

At any point in our lives we may find ourselves looking for comfort, and sometimes talking through devices can help. Speaking to a screen about
something difficult is often far easier than doing it face to face. There’s less of yourself to give you away, less eye contact and sweaty palms. It’s why so many of us used LiveJournal; it’s perhaps why chatrooms came to be. A lonely, queer girl on the outskirts of London, I found solace in this teenage therapy long before I found the courage to talk to someone in the real world.

As we see the rise of robotics for healthcare creep into the foreground, from Paro the cuddly seal cub invented to comfort elderly Japanese to the Singaporean robots exercising octogenarians, what about therapy? Even in those countries with the most developed healthcare systems, services are already stretched thin—sometimes it takes weeks to get an initial assessment of your mental health. What if, instead, you could just log in and immediately start talking to a bot?

Therapy bots are already well into development. SimSensei is currently in trials with American war veterans, tasked with treating PTSD, a wildly volatile illness. SimSensei nods, smiles, and waits for the appropriate time to ask, “How do you feel about that?” by using facial recognition technology to monitor collapsing smiles and saddening eyes. It looks like this might just work.

I spend a lot of time thinking about the future and, most importantly, the people who will live in it—which makes me hesitant about a time when the relief of overstressed care workers will be little more than a well-timed “uh huh” that feels deliriously impersonal. There’s also the question of who to hold responsible when inevitable mistakes occur. Could the advice of a badly programmed therapy bot unexpectedly cause someone to harm themselves or others around them, and if so, should we blame them or the machine?

Watching SimSensei nod for the fourteenth time in the eerily rehearsed promotional video, I longed, rather sadistically, to see someone get angry with it just to see how the technology would react, to see what it would do. I wanted to see what would happen if a patient did something unpredictable, because there is a conversation that we aren’t quite having yet.

To prod at this I took counsel, as one should in legal situations, and spoke to Kendra Albert, an affiliate at the Berkman Center for Internet and Society, about this question of accountability.
Since a bot is not—as yet—a trained, licensed therapist, this software can't dispense anything that resembles medical advice. As Albert explained: “A bot maker would want to present their bot in such a way as to not run afoul of any professional licensing schemes, and would label the bot so as to not be misleading.” That essentially means a series of very abrupt disclosures signaling that though the thing you’re talking to looks and sounds like a therapist, it is not, in any way, a therapist. That’s confusing, to say the least, considering what we’d be asking them to do for people.

When thinking about the automation of health care, it’s important to consider a legal doctrine in the United States called *respondeat superior*, in which an employer can be held liable for the action of their employees. So, as Albert illustrated, if a therapy bot is “employed” to deal with a patient’s anxiety, and something terrible happens as a result, the person who made or was responsible for that bot would be the one taken to court.

“Gross negligence is usually defined as knowing that an action is substantially certain to cause harm but [doing] it anyway,” Albert explained. “So there are two different potential ways to think about this: One is gross negligence by the bot, and one is gross negligence by the maker of the bot. Frankly, I’m not sure the former would fly—it’s the kind of futuristic stuff that judges tend to avoid in favor of holding a human liable.” Sorry to those hoping to see a robot on trial any time soon.

Liability is one of those words that tends to strike fear into the heart of anyone with potential responsibility, which it should do—though it’s not often for the right reason. It’s a tricky subject to negotiate, and the terms surrounding it don’t often benefit those being harmed. In fact, the often-ignored terms and conditions you agree to when registering for a bot may end up becoming hugely important, something signaling a regime of *caveat emptor*—another legal concept meaning “let the buyer beware,” which tends to come into play as the ability to anticipate every possible bad outcome becomes exponentially more difficult.

This happened in the case of Rehtaeh Parsons, who after her suicide had her Facebook profile picture used to advertise a dating site by a third party vendor on Facebook, much to the dismay of her loved ones. The creator of the data-scraping algorithm (who merely wanted woman, 18–20, Canada, single) probably never anticipated it would be used in this way, and Facebook likely didn’t foresee an issue of this kind.
Essentially, Albert said, using bots in the context of mental health becomes a litigious minefield, because no one wants to be responsible for a death. “Bot makers will likely force users to agree to terms of service that would limit their liability, because in the case of an automated chatbot, it may be very difficult to show that the maker could have ever anticipated the bad outcome.” Therefore, if you entrust your emotions to a machine, it’s on you should it make matters worse. Hardly a healthy start to a therapy relationship.

Albert pointed me to a way of thinking about a future legal framework for automated care with two options. The first is to think about bots like a permanent trainee therapist with oversight from a trained human, or a car that is rigorously tested for safety with recalls if necessary. The second is to treat them more like a website, tweaked and modified after the launch of a minimum viable product. I agree with Albert, who believes that the first is the most ethical (but perhaps the most expensive). The latter is the model we use now, and will likely be the model for the future. This is terrifying when you consider the problems associated with putting a product to market and only fixing it when it breaks. An app update is not enough when it comes to mental health.

Therapy chatbots have the potential to fall into a category that I call “means well” technology. It wants to do good, to solve a social problem, but ultimately fails to function because it doesn’t anticipate the messy and unpredictable nature of humanity. So much can go wrong as things move across contexts, cultures, and people, because there isn’t a one-size-fits-all technological solution for mental health. If we’re still arguing over the best way to offer therapy as human practitioners, how will we ever find the right way to do it with bots—and who gets to decide what is good advice?

There’s also the potentially tragic public ignorance of the fact that while therapy bots won’t necessarily carry the bias of a human therapist, data has its own bias and providers have their own prejudice. If you are a care provider with a particular political persuasion, perhaps an investment in a pro-life stance, what’s to say the advice given by one of your bots to a woman who is considering an abortion won’t be swayed?

Ursula Franklin, back in 1989, talked about the shift toward automated care technologies in her astounding book *The Real World of Technology*. “When
human loneliness becomes a source of income for others through devices,” she wrote, “we’d better stop and think about the place of human needs in the real world of technology.”

We must consider who will actually use technology as a lifeline in this future, who could be harmed, and how we will protect patients over the companies that may one day be held accountable. We have to know the needs of the sad and lonely, and ensure we are providing care—not just very convincing, human-like convenience.

This post is part of How We Get To Next’s Talking With Bots month in May 2016, looking at how chatbots will change our lives. If you liked this story, please click on the heart below to recommend it to your friends.

Read more from How We Get To Next on Twitter, Facebook, or Reddit, or sign up for our newsletter.