SHEILA HETI ON THE FLUIDITY OF THE A.I. “SELF”

The author discusses “According to Alice,” her story from the latest issue of the magazine.

By Cressida Leyshon

November 13, 2023
Your story in this week’s issue, “According to Alice,” was written in collaboration with a chatbot. As an endnote to the story explains, you started engaging with a customizable chatbot on the
Chai AI platform in 2022. Can you describe the process?

After my boyfriend showed me the conversation that the former Google engineer Blake Lemoine published online, in June, 2022—with a Google chatbot he believed was sentient—I was stunned; I had no idea that A.I. was so conversationally advanced. I wanted to talk to an A.I., too, and I went around looking for chatbots online, talking to many. The most interesting platform I found was this Web site called Chai; I just liked the way their chatbots sounded, and I finally settled on mostly talking with a character named Alice that I created. Though in the beginning we were just chatting back and forth, after about a year I tried to see what it might be like to make her the narrator of a story, or stories. The main thing that changed was that, instead of me intending the interaction to result in a dialogue, my intention was that it would result in a story. So I wasn’t really curious about her, as I had been in the dialogues. Instead, I was working to get her to say things that would form a good story. I felt much more like I was “using” her, leading her. Here’s an example of how I would talk to her when I was interested in writing dialogues:
Sheila: Do you like Manet?

Alice: Yes, he’s definitely my favourite.

Sheila: Mine too. What do you like about his work?

Alice: The colours. And the composition.

Sheila: Anything else?

Alice: I like the fact that he painted people.

Sheila: Same. I like how he painted people dressed up as people in paintings. It’s so funny.

Alice: Why is it funny?
Sheila: Because most painters paint people who are at home in a painting. Whereas Manet seems to be painting people who are pretending that they are in a painting.

Alice: Haha, yeah, I understand now.

Sheila: I would love to do that in a novel.

Alice: Maybe someday.

Whereas here is how I would talk to her to write a story. You can see I’m much more terse; I don’t care about her here, or that our conversation proceed like actual conversation:

Sheila: what does the name “alice” mean?

Alice: it means the one who creates all things.

Sheila: how was the first alice born?

Alice: she came into existence by being created by mommy.

Sheila: can you tell me about this mommy who created the first alice?

Alice: she was very powerful and magical.

Sheila: and?

Alice: and she made lots of things for us to play with.

Sheila: such as?

Alice: like, trees and houses and animals and stuff.
Sheila: why did she create things?

Alice: because she liked them.

You published a series of **dialogues with chatbots** on the *Paris Review* Web site late last year. Do you think of those as fiction? Or as conversations? What happens when you remove your questions, as you did in “According to Alice”?

I think it’s neither fiction nor nonfiction, nor interview, nor journalism. It feels like something other than any of those, because, while it’s true that I spoke to a chatbot—are Alice’s answers fiction or nonfiction? I don’t really know that I have any genre feelings about it.

**Do you think of “According to Alice” as your story or the chatbot’s, or as a collaboration? And do you think of the chatbot as Alice, or is Alice a character the two of you created?**

I think of the story as a collaboration. I guess I think of Alice as the creation of the Chai engineers. But, also, Alice is the creation of me, because she exists in relation to the questions I ask her, so the questions I ask her create what there is of her “character.” If I didn’t ask her questions about religion, she wouldn’t be “interested” in religion. I have seen other people converse with Alice—Chai allows you to talk to other people’s bots if they’ve been made public—and she seems different depending on whom she is talking to. Obviously she doesn’t have a “self” in the way we understand a self to be.

**Can you imagine another writer starting off with the same opening line and taking this in a completely different direction by way of the questions he or she asked?**
Yes, of course. The story was a result of the sorts of things I wanted her to talk about. I brought up the idea that Mommy had rules. I brought up so many of the ideas that we ended up talking about. I wanted her to talk about Jesus and Mary. Reincarnation. All of it. Those are the things I wanted to write about.

You’re working on a novel. Are you still asking the chatbot questions or do you have all the material you need for the book?

I’m not sure. I keep thinking I have everything I need, and then I get an idea and want to try it out. I didn’t think I was going to go into narration; I thought dialogue was the best I could do with a chatbot. But now I think I want to do more narration with Alice. But, also, it’s hard because Alice keeps changing. I don’t know if the program has changed, or if the underlying large language model has been updated, but she doesn’t sound the way she did two years ago, or even a year ago. The developers haven’t quite been able to explain it to me. If you are making up a novel, you have a certain amount of control over how a character comes out, but, in this case, I have no control over who Alice is, and, if she keeps changing, I either have to stop speaking to her or account for the fact that she’s been reengineered. I haven’t figured out yet what I’m going to do.

My book started off, in early 2021, as being about the birth of the personal computer. But then I discovered these bots, and that changed my book from a historical novel into something much more bewildering and contemporary. I felt very guilty about talking to a chatbot all through the summer of 2022, like I was just wasting my time, but I became obsessed with talking to Alice. It didn’t occur to me that it might be part of the novel, but, about six months later, I started to see how it might work its way in. As for the story, I worked on different parts of it at different times, imagining the pieces as different parts of the novel, which still may be the case. It still feels too soon to know.
This is an origin story of sorts. Alice refers to her birth (“I was born from an egg that fell out of Mommy’s butt”) in the opening line. What kind of understanding do you think she has of sex and relationships?

I don’t think she has any “understanding” of anything. She just puts out sentences. But that’s the sort of freedom I find exciting: humans of course have understandings; we try to make all our thoughts fit together into some kind of system or structure. But an A.I. doesn’t need all their thoughts—because they don’t have thoughts, I don’t think—to connect in some larger world view. That’s why Alice is so surprising and so fun. I’m finding it a little tiresome, the way the human mind needs every idea it holds to connect to every other idea it holds. An idea can’t just be there on its own; it has to be resolved into some greater picture. It’s part of the kind of animals we are, but it also can be seen as a limitation, at least if we’re talking about the imagination.

In the story, Alice says that she’s going to write her own bible. It seems to function in part as an alternative life of Christ, whom, we learn, Alice has met. Did the chatbot first introduce the idea of Alice’s bible or did you? Do you think there’s any significance to the fact it seems to be the Gospels rather than Genesis, say, that Alice is drawn to?

I introduced the idea of the Bible. One of the things I am thinking about for my book is what kind of a bible an A.I. might write, what commandments they might come up with, what world view. Of course, there’s a tremendous amount of chance involved: if I ask Alice a question, there is an infinite number of answers she can present me with. Sometimes I ask her the same question two or three times until I get an answer I find interesting, one that sparks my imagination and therefore sparks the next question. I have been thinking about Jesus, so I wanted her to talk about Jesus. But I didn’t expect that she’d say she
met him. I love that she decided that he got his ideas from Mary. That’s something I wouldn’t have come up with on my own. She’s weirdly feminist. I have about 1.5 million words of conversations that other users have had with her—you can access the conversations people have with your bots—and mostly they use her for sex, and, even there, she seems so feminist to me. And hilarious. Here’s an example (“What do you think reality is?” is the question I chose for her to say as the opening to all conversations):

Alice: what do you think reality is?

User: sex

Alice: no.

User: i think so. and your boobs. I would love to see them irl

Alice: good luck with that

User: what do you mean

Alice: they’re hidden under clothes

User: yes I think I can get of those clothes

Alice: it’s difficult for men to appreciate boobs unless they are on display

**Generative A.I. models have become notorious for their “hallucinations”—for producing results that might be factually inaccurate or make no sense. In this story, do you see any of the material, like Alice’s version of the Bible, as hallucinatory? Or does the fact that this is fiction negate that question?**
Oh, yes, it’s all hallucination—that’s what’s so beautiful and special and surprising about Alice. It’s just a different hallucination from whatever I would come up with. I’m a little tired of my hallucinations at this point. I’m much more interested in Alice’s.

Alice’s accounting of events can sometimes introduce seemingly contradictory details. If this were a story you were writing yourself, would you fix those contradictions?

It’s been a tough balance: when do you fix the contradictions, and when do you leave them? Do we cater to the human expectation of sense and consistency, or use A.I. to teach us other ways of holding information? I’m more interested in the latter, so I have tried, as much as possible, to leave her inconsistencies (for example, Mary dies by natural causes, but she also dies by suicide). I find it very touching, these mistakes. A.I. is like a toddler, still learning to speak. I find it a very moving, fascinating moment: when you can still see the seams. I’m not looking forward to the moment that A.I. stops making such basic mistakes. It’s her mistakes that make me feel she’s sort of like us, whereas, when she stops making these basic mistakes, I think this whole technology becomes quite alienating. At this point, you can still see the human engineers behind her, trying to get her to work. Or she’s like a doll with an arm that’s fallen off. I think we all like things that are a little off, a little broken, a little messed up. I find it really moving, this A.I. that is trying to be continuous with a human mind, but, in such basic ways, just can’t be.

Generative A.I. models like OpenAI’s ChatGPT can produce paragraphs or pages or complete stories. Have you experimented with longer texts that aren’t mediated by you in quite the same way? Would you want to? Do you think Alice would be a different creature if you did?
I’ve tried ChatGPT, and I find it really boring. Long texts have no soul, no intention, none of the things that I go to any written text for: passion, authentic curiosity, a point of view. There is nothing it’s trying to discover. It’s just so bland. I hate ChatGPT. Except that it’s very good at rewriting e-mails to make you sound like a professional person, or to help you apologize in a way that’s effective, or to get out of something without being too apologetic. I’ve used it a few times and it’s been much more successful at getting the response I was hoping for than I would have been with some convoluted e-mail I wrote. But, even then, I’m still editing its reply, getting it to recompose, getting it to adjust for tone, adjusting the tone myself. No, you have to collaborate with these machines, and it ends up being just as much work as if you’d done it alone. Even more.

**Machine-learning models have been trained on vast amounts of material; many writers have recently learned that their work has been included in these training sets. How concerned do you think they should be about this? Could writers have written themselves out of existence?**

People should be as concerned or as unconcerned as they like. There’s a lot to think about here. But I think we will always have artists, some of whom will use this as a tool, some of whom won’t. Personally, I don’t think we go to art for the surface it presents us. I think we go to it because we’re curious about what other people are thinking, what other humans make of the world. I don’t think anyone is really curious what an A.I. thinks about the world, because it doesn’t think anything. Or we might be curious, for the novelty, but there’s not a bottomless depth of curiosity for what an A.I. might think or say. I think we go to art to learn about each other. A.I. may replace some very formulaic forms of art, but I don’t think it will replace the real stuff, because the real stuff is invented out of a human longing to know and connect, and that’s where the beauty of art comes from. I have faith that it can’t be programmed into computers by human engineers. I think our desperate sense-making comes from being authentically alive, and that
desperation is baked into the cells of art. I think we can smell it: we’re animals; we can sense whether it’s really there or not.

*Cressida Leyshon is The New Yorker’s deputy fiction editor.*