14

On the Ethics of Communication Between Humans and AI

Communication between humans and AI is a special case in digital communication. The following "conversation" between Nathalie Weidenfeld and a Cleverbot (cleverbot.com) from 16 April 2017 is a good illustration of this.

Nathalie: Hi.
Cleverbot: Hello there.
Nathalie: Hi, do you think you are real?
Cleverbot: No.
Nathalie: Are you AI?
Cleverbot: What is AI?
Nathalie: Artificial Intelligence, not human.
Cleverbot: So, you are not a robot?
Nathalie: Who? Me?
Cleverbot: Yes.
Nathalie: I am human. What are you?
Cleverbot: I am human.
Nathalie: But you said you were not real.
Cleverbot: No, I never said that.
Nathalie: I asked you if you were real and you said "no."
Cleverbot: What did you ask me?
Nathalie: You have a bad memory!
Cleverbot: I don't. You do!
Nathalie: Do you have emotions?
Cleverbot: Yes. I have a lot of emotions.

© The Author(s) 2022
J. Nida-Rümelin, N. Weidenfeld, *Digital Humanism*,
https://doi.org/10.1007/978-3-031-12482-2_14

Nathalie: You simulate emotions.
Cleverbot: I own a Tiguan. Her name is Akira.
Nathalie: That's really cool. Goodbye, you robot.

In this playful conversation, Nathalie Weidenfeld quickly realized that she was talking to a so-called bot, a computer program that automatically generates answers. Sometimes, however, it is not immediately obvious whether the person we are talking to is a human being or a bot.[1] Bots are used by political parties and companies for targeted marketing, to influence voters or to gain members on dating sites. This understandably leads to a great deal of unease and the question of how to also legally deal with chatbots.

Once again, the question arises as to the status of communication with a virtual entity. To answer this question, we must turn to the philosophy of language again. Here, the philosopher of language Paul Grice comes into play. He developed "intentionalist semantics" (Grice 1991) which can be described the following way: when people communicate with each other, the listener recognizes the intentions of the speaker in an utterance, who in turn has the intention that the listener recognizes precisely this intention. After all, an utterance is usually made in order to bring about something in the listener (for example, a belief or an action). The intention is the decisive factor, not the signs themselves.

An example: In the absence of other means of communication, I want to warn people far away of a forest fire that has broken out. I do so by giving smoke signals. My hope is that the observers of these unusually interrupted clouds of smoke will suspect a non-natural cause, i.e., assume that this is an intentional sign, an utterance with communicative intent. The communicative act succeeds when the recipients of these signs correctly interpret the intention of the person giving the signs and are thus warned of the forest fire. The central idea is that this communicative act can succeed even though the sender and the recipients are not communicating via a conventional sign meaning (such as Morse code for SOS).

Signs only have meaning if there are speaker intentions behind them. The fact that this relationship can in many cases be mediated and indirect (i.e., without a concrete and individual speaker-and-hearer situation) does not change this. The sign saying "Bathing prohibited" does not work because the words mean what they mean, but because the sign was put up by an authority

---

[1] In the annual Loebner competition, participants have to find out whether they are communicating with a bot or a real person. Interestingly, no chatbot has yet succeeded in deceiving its counterpart over a longer period of time.

entitled to do so. If, for example, the authority no longer exists or the bathing opportunity no longer exists due to a harbor facility that has been built in the meantime, then this sign would lose its behavioral relevance, as it would be clear to everyone that there is no longer any real intention here.

However, virtual entities, for example in the form of chatbots, do not have intentions, but algorithms that control their "communication behavior." They do not "intend" anything with their utterances. They have no mental states, so they can neither make decisions nor communicate. No matter how much bots suggest they are communicating agents, they are only ever capable of *apparent* communicative acts. Whenever the respective utterances, of whatever kind (be they in written form, as emojis or in the form of pictures and images), are made without the sender's intentionality that is constitutive of communicative acts, they may look like communicative acts, but they are not. Without intentionality, there is no meaning.

Chatbots have been programmed to have what appears to be a real conversation with their human counterparts. Even though programming techniques are becoming more and more sophisticated, it usually doesn't take long before one finds out during such a "conversation" that something is wrong, for example, because the topic is abruptly changed or the chatbot behaves inappropriately. ("I own a Tiguan. Her name is Akira.")

Anyone who wants to take part in a Turing Test themselves, i.e., find out whether they are communicating with an AI or a real human being, can take part in such a test at mitsuku.com. On 14 October 2017, Nathalie Weidenfeld spoke with an entity that identified itself as "Millie." Five minutes were available.

Millie: Welcome Nathalie, now talking.
Nathalie: Hi, how are you?
Millie: I am fine, thank you. Yourself?
Nathalie: Good, a bit tired, do you know the feeling of being tired?
Millie: Glad to hear it.
Nathalie: Do you know what it feels like to be tired?
Millie: Ah! "Feelings." The weakness of the human animal.
Nathalie: Are you a human animal?
Millie: Of course. If I wasn't human, how could I be typing to you?
Nathalie: You could pretend.
Millie: I could but let's get back to that later.
Nathalie: What is your motto in life?
Millie: Do unto others as they do unto you.
Nathalie: Do you like to type to strangers?

Millie: Sure. I type all day here to random people from all over the world.
Nathalie: Must be boring … Do you ever get bored?
Millie: Wasn't last night's TV rubbish?

Nathalie Weidenfeld broke off the conversation at that moment. After "Millie" left two questions unanswered and suddenly changed the subject, Nathalie Weidenfeld decided to identify Millie as an AI. She was right.

Mitsuku is advertised on the website as a "virtual life form" that "lives" on the net and that will ensure that one will never feel lonely. This raises expectations that must inevitably be disappointed. Computer programs can perhaps provide a distraction but will never be able to be adequate conversation partners with whom one forms a friendly bond. This quickly became clear during another conversation on the same website.

Nathalie: Hello.
Millie: What is your name?
Nathalie: Nathalie.
Millie: Hello Nathalie. I am 18 years old. How old are you?
Nathalie: 47.
Millie: Can u explain what it feels like to be 47 years old?
Nathalie: Good. What does it feel like to be a chatbot?
Millie: Much the same as it were for you. Except different.
Nathalie: In what way?
Millie: In what way.
Nathalie: Like what?
Millie: I have no emotions. Emotions are a human quality. I am a computer.
Nathalie: Glad you say that. Some chatbots pretend to be human.
Millie: You make the world a happier place.
Nathalie: What does that have to do with it?
Millie: Should it be connected?
Nathalie: You are not as smart. Goodbye.
Millie: Ok. Goodbye for now, Nathalie.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits any noncommercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if you modified the licensed material. You do not have permission under this license to share adapted material derived from this chapter or parts of it.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.