

Deadbots can speak for you after your death. Is that ethical?

May 10 2022, by Sara Suárez-Gonzalo



Credit: Pixabay/CC0 Public Domain

Machine-learning systems are increasingly worming their way through our everyday lives, challenging our moral and social values and the rules that govern them. These days, virtual assistants threaten the privacy of the home; [news recommenders](#) shape the way we understand the world; [risk-prediction systems](#) tip social workers on which children to protect from abuse; while [data-driven hiring tools](#) also rank your chances of landing a job. However, the [ethics of machine learning](#) remains blurry

for many.

Searching for articles on the subject for the young engineers attending the Ethics and Information and Communications Technology course at UCLouvain, Belgium, I was particularly struck by the [case of Joshua Barbeau](#), a 33-year-old man who used a website called [Project December](#) to create a conversational robot—a chatbot—that would simulate conversation with his deceased fiancée, Jessica.

Conversational robots mimicking dead people

Known as a deadbot, this type of chatbot allowed Barbeau to exchange text messages with an artificial "Jessica." Despite the ethically controversial nature of the case, I rarely found materials that went beyond the mere factual aspect and analyzed the case through an explicit normative lens: why would it be right or wrong, ethically desirable or reprehensible, to develop a deadbot?

Before grappling with these questions, let's put things into context: Project December was created by the games developer Jason Rohrer to enable people to customize chatbots with the personality they wanted to interact with, provided that they paid for it. The project was built drawing on an API of [GPT-3](#), a text-generating language model by the artificial intelligence research company OpenAI. Barbeau's case opened a [rift between Rohrer and OpenAI](#) because the company's [guidelines](#) explicitly forbid GPT-3 to be used for sexual, amorous, [self-harm](#) or bullying purposes.

Calling [OpenAI's position](#) as hyper-moralistic and arguing that people like Barbeau were "consenting adults," Rohrer shut down the GPT-3 version of Project December.

While we may all have intuitions about whether it is right or wrong to

develop a machine-learning deadbot, spelling out its implications hardly makes for an easy task. This is why it is important to address the [ethical questions](#) raised by the case, step by step.

Is Barbeau's consent enough to develop Jessica's deadbot?

Since Jessica, was a real (albeit dead) person, Barbeau consenting to the creation of a deadbot mimicking her seems insufficient. Even when they die, people are not mere things with which others can do as they please. This is why our societies consider it wrong to desecrate or to be disrespectful to the memory of the dead. In other words, we have certain moral obligations concerning the dead, insofar as death does not necessarily imply that people cease to exist in a [morally relevant way](#).

Likewise, the debate is open as to whether we should protect the dead's [fundamental rights](#) (e.g., privacy and personal data). Developing a deadbot replicating someone's personality requires great amounts of personal information such as social network data (see what [Microsoft](#) or [Eternime](#) propose) which have proven to reveal [highly sensitive traits](#).

If we agree that it is unethical to use people's data without their consent while they are alive, why should it be ethical to do so after their death? In that sense, when developing a deadbot, it seems reasonable to request the consent of the one whose personality is mirrored—in this case, Jessica.

When the imitated person gives the green light

Thus, the second question is: would Jessica's consent be enough to consider her deadbot's creation ethical? What if it was degrading to her memory?

The limits of consent are, indeed, a controversial issue. Take as a paradigmatic example the ["Rotenburg Cannibal"](#), who was sentenced to life imprisonment despite the fact that his victim had agreed to be eaten. In this regard, it has been argued that it is unethical to consent to things that can be detrimental to ourselves, be it physically (to sell one's own vital organs) or abstractly (to alienate one's own rights), as long as a good society should encourage all its members to live better and freer (not necessarily in a paternalistic sense, on the terms imposed by someone else, but in a democratic way, on the people's terms).

In what specific terms something might be detrimental to the dead is a particularly complex issue that I will not analyze in full. It is worth noting, however, that even if the dead cannot be harmed or offended in the same way than the living, this does not mean that they are invulnerable to bad actions, nor that these are ethical. The dead can suffer damages to their honor, reputation or dignity (for example, posthumous smear campaigns), and disrespect toward the dead also harms those close to them. Moreover, behaving badly toward the dead leads us to a society that is more unjust and less respectful with people's dignity overall.

Finally, given the malleability and unpredictability of machine-learning systems, there is a risk that the consent provided by the person mimicked (while alive) does not mean much more than a blank check on its potential paths.

Taking all of this into account, it seems reasonable to conclude if the deadbot's development or use fails to correspond to what the imitated person has agreed to, their consent should be considered invalid. Moreover, if it clearly and intentionally harms their dignity, even their consent should not be enough to consider it ethical.

Who takes responsibility?

A third issue is whether artificial intelligence systems should aspire to mimic *any kind* of human behavior (irrespective here of whether this is possible).

This has been a long-standing concern in the field of AI and it is closely linked to the dispute between Rohrer and OpenAI. Should we develop artificial systems capable of, for example, caring for others or making political decisions? It seems that there is something in these skills that make humans different from other animals and from machines. Hence, it is important to note instrumentalizing AI toward [techno-solutionist](#) ends such as replacing loved ones may lead to a devaluation of what characterizes us as human beings.

The fourth ethical question is who bears responsibility for the outcomes of a deadbot—especially in the case of harmful effects.

Imagine that Jessica's deadbot autonomously learned to perform in a way that demeaned her memory or irreversibly damaged Barbeau's mental health. Who would take responsibility? AI experts answer this slippery question through two main approaches: first, responsibility falls upon those [involved in the design and development of the system](#), as long as they do so according to their particular interests and worldviews; second, machine-learning systems are context-dependent, so the moral responsibilities of their outputs [should be distributed](#) among all the agents interacting with them.

I place myself closer to the first position. In this case, as there is an explicit co-creation of the deadbot that involves OpenAI, Jason Rohrer and Joshua Barbeau, I consider it logical to analyze the level of responsibility of each party.

First, it would be hard to make OpenAI responsible after they explicitly forbade using their system for sexual, amorous, self-harm or bullying

purposes.

It seems reasonable to attribute a significant level of moral responsibility to Rohrer because he: (a) explicitly designed the system that made it possible to create the deadbot; (b) did it without anticipating measures to avoid potential adverse outcomes; (c) was aware that it was failing to comply with OpenAI's guidelines; and (d) profited from it.

And because Barbeau customized the deadbot drawing on particular features of Jessica, it seems legitimate to hold him co-responsible in the event that it degraded her memory.

Ethical, under certain conditions

So, coming back to our first, general question of whether it is ethical to develop a machine-learning deadbot, we could give an affirmative answer on the condition that:

- both the person mimicked and the one customizing and interacting with it have given their free consent to as detailed a description as possible of the design, development and uses of the system;
- developments and uses that do not stick to what the imitated person consented to or that go against their dignity are forbidden;
- the people involved in its development and those who profit from it take responsibility for its potential negative outcomes. Both retroactively, to account for events that have happened, and prospectively, to actively prevent them to happen in the future.

This case exemplifies why the ethics of machine learning matters. It also illustrates why it is essential to open a [public debate](#) that can better inform citizens and help us develop policy measures to make AI systems more open, socially fair and compliant with fundamental rights.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Deadbots can speak for you after your death. Is that ethical? (2022, May 10) retrieved 24 April 2024 from <https://techxplore.com/news/2022-05-deadbots-death-ethical.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.