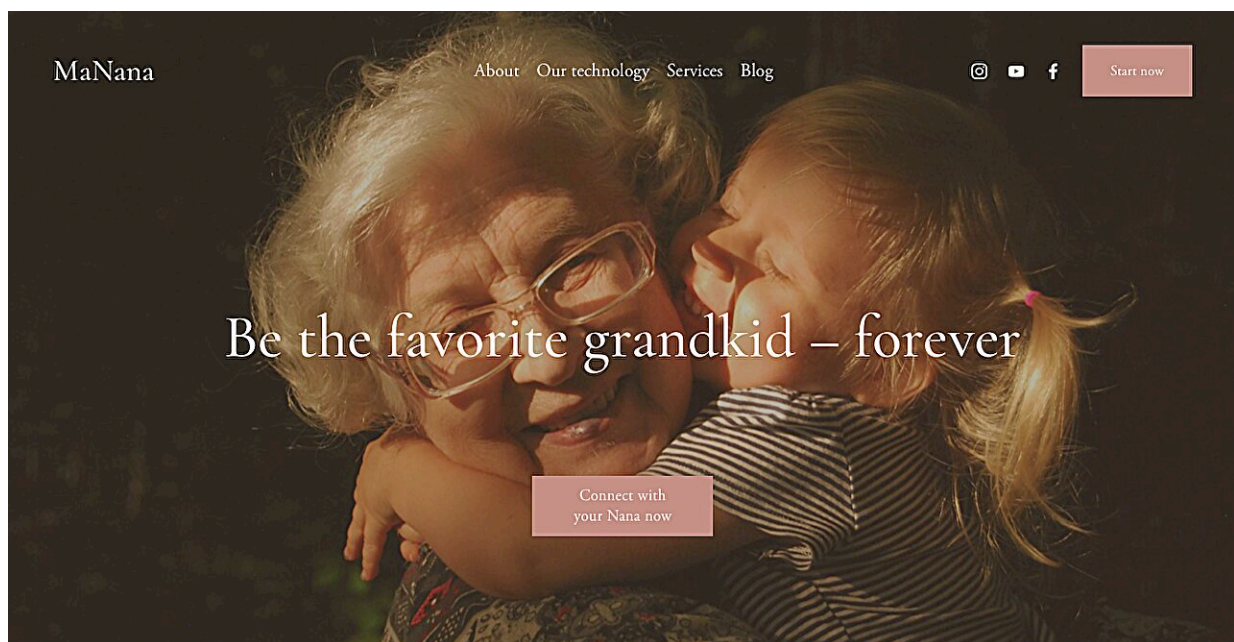# 'Digital afterlife': Call for safeguards to prevent unwanted 'hauntings' by AI chatbots of dead loved ones

May 8 2024

A visualisation of a fictional company called MaNana, one of the design scenarios used in the paper to illustrate the potential ethical issues in the emerging digital afterlife industry. Credit: Dr. Tomasz Hollanek

Without design safety standards, artificial intelligence that allows users to hold text and voice conversations with lost loved ones runs the risk of causing psychological harm and even digitally "haunting" those left

behind, according to University of Cambridge researchers.

"Deadbots" or "Griefbots" are AI chatbots that simulate the language patterns and personality traits of the dead using the digital footprints they leave behind. Some companies are already offering these services, providing an entirely new type of "postmortem presence."

AI ethicists from Cambridge's Leverhulme Centre for the Future of Intelligence outline three design scenarios for platforms that could emerge as part of the developing "digital afterlife industry," to show the potential consequences of careless design in an area of AI they describe as "high risk."

The research, published in the journal *Philosophy and Technology*, highlights the potential for companies to use deadbots to surreptitiously advertise products to users in the manner of a departed loved one, or distress children by insisting a dead parent is still "with you."

When the living sign up to be virtually recreated after they die, resulting chatbots could be used by companies to spam surviving family and friends with unsolicited notifications, reminders and updates about the services they provide—akin to being digitally "stalked by the dead."

Even those who take initial comfort from a deadbot may get drained by daily interactions that become an "overwhelming emotional weight," argue researchers, yet may also be powerless to have an AI simulation suspended if their now-deceased loved one signed a lengthy contract with a digital afterlife service.

"Rapid advancements in generative AI mean that nearly anyone with Internet access and some basic know-how can revive a deceased loved one," said Dr. Katarzyna Nowaczyk-Basińska, study co-author and researcher at Cambridge's Leverhulme Centre for the Future of

Intelligence (LCFI).

"This area of AI is an ethical minefield. It's important to prioritize the dignity of the deceased, and ensure that this isn't encroached on by financial motives of digital afterlife services, for example. At the same time, a person may leave an AI simulation as a farewell gift for loved ones who are not prepared to process their grief in this manner. The rights of both data donors and those who interact with AI afterlife services should be equally safeguarded."
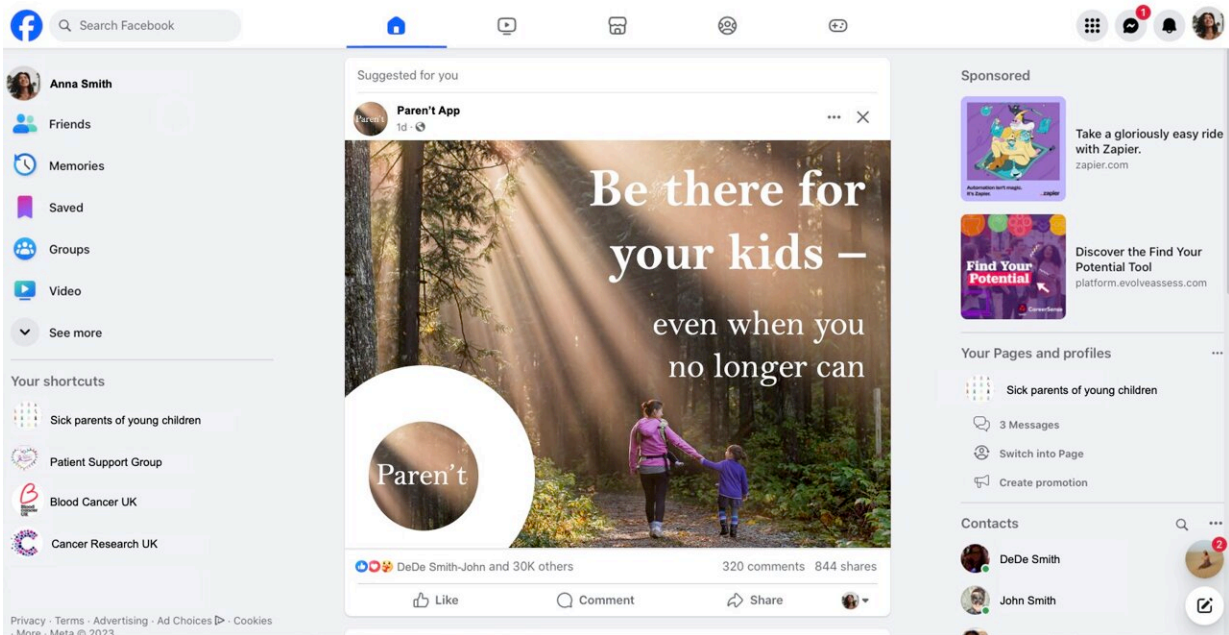
A visualisation of a fictional company called MaNana, one of the design scenarios used in the paper to illustrate the potential ethical issues in the emerging digital afterlife industry. Credit: Dr. Tomasz Hollanek

Platforms offering to recreate the dead with AI for a small fee already exist, such as "Project December," which started out harnessing GPT

models before developing its own systems, and apps including "HereAfter." Similar services have also begun to emerge in China.

One of the potential scenarios in the new paper is "MaNana," a conversational AI service allowing people to create a deadbot simulating their deceased grandmother without consent of the "data donor" (the dead grandparent).

The hypothetical scenario sees an adult grandchild who is initially impressed and comforted by the technology start to receive advertisements once a "premium trial" finishes. For example, the chatbot suggesting ordering from food delivery services in the voice and style of the deceased.

The relative feels they have disrespected the memory of their grandmother, and wishes to have the deadbot turned off, but in a meaningful way—something the service providers haven't considered.

"People might develop strong emotional bonds with such simulations, which will make them particularly vulnerable to manipulation," said co-author Dr. Tomasz Hollanek, also from Cambridge's LCFI.

"Methods and even rituals for retiring deadbots in a dignified way should be considered. This may mean a form of digital funeral, for example, or other types of ceremony depending on the social context. We recommend design protocols that prevent deadbots being utilized in disrespectful ways, such as for advertising or having an active presence on social media."

While Hollanek and Nowaczyk-Basińska say that designers of recreation services should actively seek consent from data donors before they pass, they argue that a ban on deadbots based on non-consenting donors would be unfeasible.

They suggest that design processes should involve a series of prompts for those looking to "resurrect" their loved ones, such as "Have you ever spoken with X about how they would like to be remembered?" so the dignity of the departed is foregrounded in deadbot development.
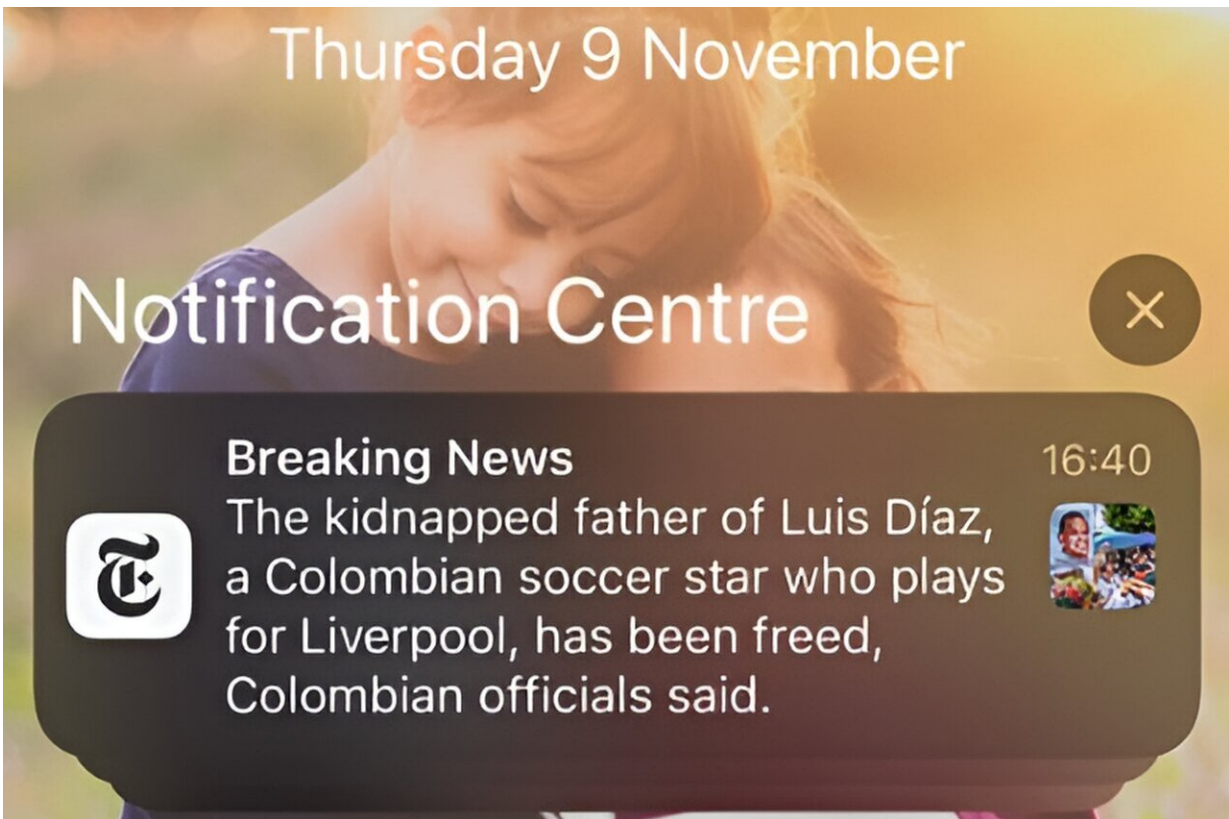
A visualisation of a fictional company called Stay, one of the design scenarios used in the paper to illustrate the potential ethical issues in the emerging digital afterlife industry. Credit: Dr. Tomasz Hollanek

Another scenario featured in the paper, an imagined company called "Paren't," highlights the example of a terminally ill woman leaving a deadbot to assist her eight-year-old son with the grieving process.

While the deadbot initially helps as a therapeutic aid, the AI starts to generate confusing responses as it adapts to the needs of the child, such as depicting an impending in-person encounter.

The researchers recommend age restrictions for deadbots, and also call for "meaningful transparency" to ensure users are consistently aware that they are interacting with an AI. These could be similar to current warnings on content that may cause seizures, for example.

The final scenario explored by the study—a fictional company called "Stay"—shows an older person secretly committing to a deadbot of themselves and paying for a twenty-year subscription, in the hopes it will comfort their adult children and allow their grandchildren to know them.

After death, the service kicks in. One adult child does not engage, and receives a barrage of emails in the voice of their dead parent. Another does, but ends up emotionally exhausted and wracked with guilt over the fate of the deadbot. Yet suspending the deadbot would violate the terms of the contract their parent signed with the service company.

"It is vital that digital afterlife services consider the rights and consent not just of those they recreate, but those who will have to interact with the simulations," said Hollanek. "These services run the risk of causing huge distress to people if they are subjected to unwanted digital hauntings from alarmingly accurate AI recreations of those they have lost. The potential psychological effect, particularly at an already difficult time, could be devastating."

The researchers call for design teams to prioritize opt-out protocols that allow potential users terminate their relationships with deadbots in ways that provide emotional closure.

Nowaczyk-Basińska added, "We need to start thinking now about how

we mitigate the social and psychological risks of digital immortality, because the technology is already here."

More information: Griefbots, Deadbots, Postmortem Avatars: on Responsible Applications of Generative AI in the Digital Afterlife Industry, *Philosophy & Technology* (2024). [DOI: 10.1007/s13347-024-00744-w](#)

Provided by University of Cambridge