

THE ETHICS OF AI-POWERED VOICE CHATBOTS AND VIRTUAL REALITY FOR DIGITAL AFTERLIVES

Eun Chang Choi
Impactive AI
28, Seolleung-ro 92-gil, Gangnam-gu
Seoul, South Korea
E-mail: eunchang.choi@aya.yale.edu

We have been facing an ethical and legal vacuum that needs to be filled in terms of using data of the deceased afterlife. Ethical aspects of the digital afterlife service are largely unexplored. The deep learning powered re-animated images and voice chatting with the deceased in virtual reality sphere are bringing psychological effect of being re-united. AI-powered virtual reality or metaverse will be a space where the bereaved could meet resurrected persons. It is highly likely that the concepts of digital grief would become a new culture. Despite of the rise of digital afterlife industry, virtual immortality raises challenging ontological questions about the demarcation between life and death.

1. The Quest for Virtual Immortality

More than three decades ago, a renowned roboticist had predicted that “More radically, we could *download* our minds directly into a body in the simulation and *upload* back into the real world when our mission is accomplished” in his book *Mind Children*.¹ Immortality, Hans Moravec suggested, is only a temporary defense against the wanton loss of knowledge and function that is the worst aspect of personal death. In order to achieve human equivalence, robots are expected to have the capacity to perform ten trillion calculations per second. The machines will have capacity to reason and ability to perceive, interact with complex environment. Most of all, the ability to reproduce biological humans into machines with minds is a sine qua non for the members of any race that aspires to immortality, he envisaged. But, his idea about ‘personal immortality by mind transplant’ has not been realized.

Moravec’s controversial quest for immortality is not anymore science fantasy or a mind-boggling glimpse in virtual world. Meanwhile, robotic engineering technology has yet reached to point in which downloads human mind into machines. But digital data the dead left can be transformed into *digital incarnation* which represents *another self* of the dead.

When it comes to uploading minds of the deceased, ‘digital remains’— the artefacts and data users consciously or unconsciously produced using email accounts, texts, social network services and other online instant messaging apps like WhatsApp, WeChat, Kakao Talk, Facebook Messenger etc.—have become increasingly available.² Various personal data can be used for replicating his/her social behaviour and the habit of using words in talks.

For instance, an episode of British science fiction *Black Mirror* has a storyline that a company created a new virtual character with dead husband’s data of online communications and social media profiles.³ The virtual character can chat with his wife in real time basis. She decided to upload more videos and photos of the husband to talk with the artificial character over the phone. ‘MyHeritage’ or ‘Deep Nostalgia’ sites allows users to produce a remarkably convincing short video with old photos of a long-lost relative or friend they uploaded. The reanimated reality can be used for family reunion as well as for satire, mockery or entertainment. A chatbot *Replika* could reproduce users identity by learning real persons’ communication style in preparation of users’ death.

Öhman & Florid suggested four types of categorization in posthumous online presence that *Digital Afterlife Industry(DAI)* offers: i) Information management service, ii) Online memorial services, iii) Posthumouse messaging service, and iv) Re-creation service. In practice,

posthumouse messaging is carried out by communication tool. And re-animation is performed by artificial agents and virtual reality technology.⁴

This paper will investigate the ethical questions focusing iii) and iv). Numerous startups are already anticipating growing demand on digital mourning. In addition, online platforms like Facebook and Google have developed features which algorithms curate digital artefacts taken from personal online archives. What is more, Microsoft has been granted a patent that aims to revive the deceased as an artificial intelligence chatbot by preserving historical figures and living people.⁵ But ethical aspects of data use and AI chatbots in digital afterlife service are largely unexplored.

2. Family Re-Union in Virtual World



Digitally resurrected persona of the dead is blurring the distinction between the quick and the dead. In the Korean documentary, entitled *Meeting You*, aired out in 2020, deep learning combined with virtual reality allowed a grieving mother to reunite with her seven-year-old-daughter, who died from blood-related diseases.⁶ Digital remains of the deceased daughter included videos footages, daily conversations, family photos, preference of food, dressing style, and memory of places.

Virtual reality offered the mother with immersive experience that is different from common behaviors of mourning such as looking at photos, videos or un-animated avatars, which are incapable of interactive communications. More importantly, the mother touched her daughter with haptic gloves and head mounted display. The late daughter continuously chats with her mother on face-to-face. They were singing happy birthday song together. The mother kept weeping racked with emotion but she admitted that her daughter is dead.

Chatbots and re-animations provided the bereaved with chance to share feelings about their loved ones. Virtual experience is proved to be an effective tool of consolation and new therapeutic application on behalf of those who had lost loved ones in grief experience.⁷ With the possible long-term psychological impacts of virtual gathering on users is unclear, the mother described the re-union with re-animated daughter as satisfactory experience even though it was a very brief.

3. The Ethics of Chatting With the Dead

At a glance, chatting with the dead seems scary. But bereaved family wants to snuggle up to the lost member in virtual reality world letting alone chatting. In brief, perspectives are different. A recently developed chatbot *Replika* (<https://replika.ai/>) asks a series of questions to users, so that it eventually learns to mimic users' communication style.⁸ *Replika* has been developed to give an actual voice to the digital ghosts of the deceased—seamless virtual communication even afterlife. The chatbot also has sufficient emotional responsiveness that makes users feel *Replika* as more of a virtual friend. At the end of day, the capability of *Replika* could reproduce users presence once they are dead. It deployed sequence-to-sequence methods to learn how to think and speak like a human by processing transcripts of conversations they've had during their lifetimes. Many *Replika* users confided in regardless of privacy. The capability of *Replika* makes it way more difficult to discern the voice to the digital chatbot of the deceased and real persons alive.

The first chatbot program inventor Joseph Weizenbaum raised concern over too much trust on chatbot program ELIZA.⁹ It was an early natural language processing computer program created in 1960s at MIT Artificial Intelligence Lab. At the time, Weizenbaum noticed that many users formed emotional attachments to the algorithm through their interactions with Eliza. Eliza chatbot demonstrated the tendency of users who were 'willingly' or even 'complacently' falling into the illusion that conversation with the chatbot was real relationship. Weizenbaum's concern has been based on the disparity between overwhelming trust and the simplicity of Eliza design that has been simulated a Rogerian psychotherapist. Eliza became a veritable archetype of the deceitful character of AI.¹⁰ (Natale 2021)

4. Chatbot Anthropomorphism

Chatbot anthropomorphism means the misperception that users tend to consider AI-powered voice chatbot as a real person with sentience. AI-powered voice chatting gets closer to conversational with real persons. Therefore, voice chatting with the deceased loved ones would amplify emotional empathy with a strong attachment to virtually resurrected digital persona. Then voice chatbots may be evolved beyond preserving a personality of the dead. For instance, users willingly reckon that the voice of chatbot systems could be given a legal personhood as if the deceased has never died.

Should Weizenbaum's concern be applicable to conversation via advanced chatbots with the dead? Human-extended machine cognition is considered to be a specific form of artificial intelligence. In addition, it has been also suggested that language is the basis for the brain to carry out cognition. The voice chatbots are not a system that mastered language senses or grasp the context of conversation with carefully refined words. But users unwittingly expect a higher capacity in machine judgments and communication skills (Salles et al. 2020).¹¹ Although voice chatbots appear to make an automated conversation, they only carry out tasks selecting one of lines in corpus datasets as designed by a human for a specific purpose with limited capabilities.

But technological settings and the purpose of chatting with the deceased are different from Eliza in 1960s. The goal of chatting with the dead in a way of digital mourning is to remember them and to get the feeling close presence of them rather than unconditional positive regard or counseling. In business interactions, chatbot anthropomorphism tends to invoke a negative effect on customer satisfaction that a limited conversation puts customers into angry emotional states.¹² But in occasion of digital mourning, there would be not much room that chatbot anthropomorphism could harm or deceive users. As far as chatting conversations are being carried out with a view to consoling and psychological effect of being re-united with the

deceased family members or friends, it would be not deceitful unless conversation threads are compromised by unexpected the third party's intervention or imposter identity of the dead.

In technological perspective, seamless conversation between users and AI-powered voice chatbots resembling the deceased is the outcome of human-machine-symbiosis research. In other words, the close coupling of the deceased and voice chatbots in virtual reality indicates a move to *extended mind*.¹³ It seems like a quasi human-machine symbiosis that the dead human's data cooperates with chatbot system. Unfortunately, Weizenbaum's concern on chatbot anthropomorphism was contrasted with the vision of human-machine symbiosis.

5. Data Consent for Afterlife

When digital mourning service and virtual re-union becomes common culture, its service architects will be inevitably relying on data of the deceased. Therefore, permanent deletion of social media data or Internet service account will deprive of opportunity to build a digital memorial service or virtual immortality that bridge the gap between digital afterlife and physical world.

But not many individuals are familiar with terms of service with regard to data after death. The world-largest social media seems to have acknowledged the claim to control data afterlife. Like Facebook. Instagram's data policy offers a memorialization option but it requires proof of death such as obituary or news. Facebook, Inc., has put in place measures to control the post-mortem data on their site. The company publicly noted its afterlife policy as follows: "Memorialized accounts are a place for friends and family to gather and share memories after a person has passed away." "Content the person shared (example: photos, posts) stays on Facebook and is visible on Facebook to the audience it was shared with."¹⁴ Otherwise, Facebook users could choose to have their account with data permanently deleted after they pass away. Verified family members also may request the removal of a loved one's account. Memorialized profiles don't appear in public spaces such as in suggestions for 'People You May Know', ads or birthday reminders.

Google's Inactive Account Manager also lets users make a plan their own digital afterlife. Users can delete Gmail messages, Blogger, Google Voice, YouTube and Google+ profile as a digital inheritance. Users can choose specific person who can control their digital assets and which of them would like to share personal data with.¹⁵ Twitter accepts reporting the death of a user and choose between a number of options for what to do with the deceased person's account. Apple' iPhone users with iOS 15.2 operating system could indicate who can directly inherit their data after death. Apple also allows family access to iPhone data.

Unlike Facebook or Google, Naver Corp., a South Korea's Internet giant, considers user account and password as confidential data then it won't provide it to the bereaved families even if requested. It is because passwords are encrypted in a way that cannot be decrypted. Thus, it only offers backup services for publicly available data to the bereaved families due to the concern over privacy¹⁶.

According to current laws in South Korea, personal data itself is not a property that should be inherited to immediate family. In a similar vein, the European General Data Protection Regulation no longer applies to identifiable data that relate to a personal data of the deceased.¹⁷ Currently the control of data in social media is being handled by platform's afterlife data policy. But in case of the absence of afterlife data policy or data ethics guideline, legal disputes will arise. Prior to the dispute occurring, this ethical and legal vacuum needs to be filled regarding data ownership of the deceased. For instance, it would be ideal to set a default rule on 'prior permission ractice'. That is, anyone should set limits or conditions regarding the use of his/her own data and digital remains for afterlife, so that unauthorized third party cannot create virtual human equivalence of specific person without prior informed consent of himself/herself or ex post permission from the bereaved family.

6. Post-human Harms in Non-consensual Use

With regard to negative use of deepfakes using an AI algorithm, concerns have been raised about blackmail, intimidation, ideological manipulation, accountability and intimidation. There is growing public outcry over deepfake tools that swap real females faces – celebrities and South Korean K-Pop stars into characters of porn. AI-generated synthetic media tools and apps could strip clothes off any targeted woman photographed and create pornographic representations in a few minutes.¹⁸ A study found that the prominence of non-consensual deepfake pornography, which accounted for 96% of the total deepfake videos online.¹⁹

In addition to the deepfake pornography, deepfakes are also being used in politics. In early 2022, a political party in South Korea uploaded a deepfake video of a former President Roh to Youtube platform without permission of Roh's family. The video contained AI-powered voice of Roh and it was regarded as an attempt to revive the late political icon. Roh's voice portrayed an endorsement for a presidential candidate of a ruling party. But the political claim with virtual narrative got backfires from the public. The virtual endorsement by a former President Roh was regarded as illegitimate practice consisting of defamation of the deceased.²⁰ Despite of Roh's death, his political power remains palpable in the political sphere. And it cannot be political propaganda tool or legacy property of a particular political party to which he had once been affiliated. The happening connoted technological advancement in the manipulation of political figures.

Even though deepfake technology isn't yet enough to completely deceive viewers, it presents significant ethical challenges, in particular when data of the deceased used without permission of the bereaved or beyond presumptive intention of the deceased. It is noteworthy that even posthumous events occur after death could harm the deceased. Some argue, however, that as long as the deceased was a public figure, there will be no room of harm in light of freedom of expression. Data of celebrities and notable figures who have passed away is not always exclusive property of their family. It is because their data—voice, photos, videos, and public speech—is already recorded online. Nonetheless but, it does not necessarily mean that harm or intentional dishonoring the post-human should be allowed. For instance, after Anthony Bourdain's death, a documentary filmmaker decided to use AI to generate the voice of the late Bourdain. The artificial voice were only three particular lines in a documentary, it was enough to provoke anger of his fans.²¹

7. The Limit of Digital Immortality

Digital Afterlife Industry(DAI) will deploy State-of-the-art AI and virtual reality for digital immortality. And they will be focusing the vivid manifestation of human equivalence in digital form. Re-creation services inevitably need a range of personal data in order to generate similar conversations resembling the way in which the dead spoke it. Typically, a voice chatbot could pronounce new messages based on past data.²² At the same time, it would raise ontological questions about the demarcation line between life and death.

Savin-Baden & Burden argues that artificial general intelligence-based virtual human could gain its own sentience and consciousness. Their viewpoints presume that coding a digital immortal would create sentience of it (Maggi Savin-Baden & David Burden 2019).²³ The possibility of autonomous presence could make us virtual sapiens. However, it is too early to predict the feasibility of everlasting digital immortality since it hinges on the advanced Human-level AI. In the long run, deep learning technology focused on constructing virtual reality could confer the manifestation or semblance of digital immortality, but digital personas in virtual reality could not have their own sentience in virtual form.

Will should be lodged with lawyers during lifetime. Except for 5~7 percentage in South Korea, the majority of people fails to make their will before death. More than half of American adults do not leave a will. As a consequence, family members and business partners are supposed to be embroiled in complicated legal disputes. The tidal wave of resurrecting

digital person afterlife would increase the expectation that the remark of virtual human equivalence could be interpreted as presumptive intention of the deceased who left no will.

References

1. Moravec, Hans, *Mind Children: The Future of Robot and Human Intelligence* (1988)
2. Lambert, Alex et al., Algorithmic memorial videos: Contextualising automated curation, *Memory Studies* Vol. 11 Issue. 2, (Nov. 2016)
3. Brooker, C. & Harris, O. *Be right back* In *Black Mirror*, Channel 4, UK (2013)
4. Öhman, C. & Floridi, L. The Political Economy of Death in the Age of Information: A Critical Approach to the Digital Afterlife Industry. *Minds Mach.* (2017)
5. U. S Patent, 'Creating a conversational chat bot of a specific person' US10853717B2.
6. VR Human Documentary, *Meeting You*, MBC
http://content.mbc.co.kr/program/documentary/3479845_64342.html
7. Silvia Francesca Maria Pizzoli et al., From virtual to real healing: a critical overview of the therapeutic use of virtual reality to cope with mourning, *Curr Psychol.* (Aug. 2021)
8. Davide Sisto, *Online Afterlives: Immortality, Memory, and Grief in Digital Culture*, MIT Press (2020)
9. Joseph Weizenbaum, *Computer Power and Human Reasons* (1976)
10. Simone Natale, *Deceitful Media: Artificial Intelligence and Social Life* (2021)
11. Arleen Salles, Kathinka Evers & Michele Farisco, Anthropomorphism in AI, *AJOB Neuroscience*, 11:2, 88-95 (2020)
12. Cammy Crollic et al., Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions, *Journal of Marketing* (2021)
13. Andy Clark, Language, Embodiment, and the Cognitive Niche, *Trends in Cognitive Sciences* 10, no. 8 (2006)
14. Facebook, What will happen to my Facebook account if I pass away?
<https://www.facebook.com/help/103897939701143> (Visited. April 2022)
15. Google Inactive Account Manager
<https://support.google.com/accounts/answer/3036546?hl=en>
16. Digital Remains and Privacy Policy after Death, Naver <https://bit.ly/39kwPoV>
17. What is considered personal data under the EU GDPR? <https://gdpr.eu/eu-gdpr-personal-data/>
18. Karen Hao, A horrifying new AI app swaps women into porn videos with a click, *MIT Technology Review* (2021)
19. Henry Ajder, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen, The State of Deepfakes: Landscape, Threats, and Impact (2019)
20. Ruling party's attempt to revive late progressive icons backfires, *Korean Herald* (Feb. 7 2022)
21. Helen Rosner, The Ethics of a Deepfake Anthony Bourdain Voice, *The New Yorker* (July 2021)
22. Carl J Öhman, Are the dead taking over Facebook? A Big Data approach to the future of death online, *Big Data & Society* (2019)
23. Maggi Savin-Baden & David Burden, Digital Immortality and Virtual Humans, *Postdigital Science and Education*. Vol. 1, 87–103 (2019)