Generative Ghosts: Anticipating Benefits and Risks of AI Afterlives

MEREDITH RINGEL MORRIS, Google DeepMind, USA
JED R. BRUBAKER, University of Colorado Boulder, USA

As AI systems quickly improve in both breadth and depth of performance, they lend themselves to creating increasingly powerful and realistic agents, including the possibility of agents modeled on specific people. We anticipate that within our lifetimes it may become common practice for people to create custom AI agents to interact with loved ones and/or the broader world after death; indeed, the past year has seen a boom in startups purporting to offer such services. We call these *generative ghosts*, since such agents will be capable of generating novel content rather than merely parroting content produced by their creator while living. In this paper, we reflect on the history of technologies for AI afterlives, including current early attempts by individual enthusiasts and startup companies to create generative ghosts. We then introduce a novel design space detailing potential implementations of generative ghosts, and use this analytic framework to ground discussion of the practical and ethical implications of various approaches to designing generative ghosts, including potential positive and negative impacts on individuals and society. Based on these considerations, we lay out a research agenda for the AI and HCI research communities to better understand the risk/benefit landscape of this novel technology so as to ultimately empower people who wish to create and interact with AI afterlives to do so in a beneficial manner.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence**; • **Human-centered computing** → **Human computer interaction (HCI)**.

Additional Key Words and Phrases: AI, AI agents, Generative AI, HCI, digital afterlife, digital legacy, post-mortem AI, post-mortem data management, end-of-life planning, death, griefbots

ACM Reference Format:
Meredith Ringel Morris and Jed R. Brubaker. 2018. Generative Ghosts: Anticipating Benefits and Risks of AI Afterlives. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 23 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The past few years have brought incredible growth in the capabilities of generative AI models, particularly large language models (LLMs) such as GPT-4 [86], Palm 2 [3], and Llama 2 [112], though there has also been incredible progress in generative AI for the production of images [93], video [103], and audio [10], as well as a new generation of multimodal models [36, 121] that combine functionality across several media categories. These models, in turn, have given rise to new types of *generative agents* [88], simulacra that can produce believable human behaviors, including capabilities such as memory and planning. While still in their infancy, generative agents and related technologies are likely to increase in fidelity and popularity as underlying model capabilities improve, compute costs drop, and the required technical expertise decreases. For instance, in November 2023 OpenAI released GPTs [87], a no-code interface

Authors' Contact Information: Meredith Ringel Morris, merrie@google.com, Google DeepMind, Seattle, Washington, USA; Jed R. Brubaker, jed.brubaker@colorado.edu, University of Colorado Boulder, Boulder, Colorado, USA.

Manuscript submitted to ACM

for people to develop agentic AIs. As AI models increase the set of human capabilities they can faithfully reproduce [18, 75], societal change is inevitable. For instance, experts anticipate that powerful AI systems may profoundly change disparate areas of society such as the labor market [27], the education system [52], the pursuit of scientific knowledge [74], and criminal activities [30]. In this paper, we discuss how advances in AI might change personal and cultural practices around death and dying as well.

We introduce the concept of *generative ghosts*[1], AI agents that represent a deceased person, and discuss why we anticipate such representations may become increasingly common. We share examples from recent media reports of individuals and startups experimenting with the creation of generative ghosts, a trend that seems to be accelerating, particularly in East Asia. We introduce a design space to describe possible instantiations of generative ghosts and use this analytic framework to explore how design choices might lead to practical and ethical concerns as well as potentially beneficial outcomes.

Our contributions include: (1) identifying and characterizing an emerging phenomenon of creating "generative ghosts" to represent the deceased, (2) introducing a design space of dimensions and analysis of potential benefits and harms that can be used to support future empirical research and motivate fieldwork. By characterizing this emerging trend, highlighting potential risks, and creating a framework for future investigation, we aim to ensure that future technical and sociotechnical systems will maximize the potential benefits of "AI afterlives" while minimizing potential risks.

2 Related Work

We discuss the rich literature on how technologies have changed practices around death and dying and initial forays into AI afterlives by individuals and start-up ventures.

2.1 Post-Mortem Technology

Throughout history, people have turned to technology to remember, memorialize, and even interact with the dead. Gravestones and other burial markers can be traced nearly back to 3000 B.C.E. [110]. Obituaries in the U.S., while dating back to the 16th century, became more common during the 19th century in part due to the U.S. Civil War [47] – an event that also brought embalming into favor. Even the mediums of the Spiritualism movement in the late 19th and early 20th century turned to telegraphs, radio-wave detectors, and later wireless radio in their attempts to detect the presence of and communicate with the dead [79].

During the earliest days of the World Wide Web, when people would create personal Home Pages describing their lives and family, it was routine for people to dedicate a page to the memory of a family member, often a deceased parent or family pet [11]. Online graveyards, websites specifically dedicated to the memorialization of the dead, soon followed [95, 96]. The *Virtual Memorial Garden* is the earliest documented example [96]. Created in 1996, it featured a collection of brief obituaries authored by loved ones, capturing their grief, the circumstances around the death, and loved ones' guilt.

Even as people adopted digital technology to memorialize the dead, scholars in human-computer interaction and social computing began to note how mortality is overlooked by technology design and have argued for increased attention to end-of-life and mortality [8, 67, 68, 85], often termed "thanatosensitive design" [67]. Scholars have studied

[1]We use the term "ghosts" to encapsulate both the memorializing practices we imagine these agents may fulfill and the unintentional persistent presences we anticipate such AI agents will bring.

the intersection of technology and mortality across diverse contexts, including digital heirlooms (e.g., [9, 84]), communal rituals (e.g., [48, 113]), online memorials (e.g., [41, 66, 72]), digital legacy (e.g., [25, 39–41]), and family archives (e.g., [53]). Yet, as Doyle and Brubaker summarize [26], end-of-life scenarios also present new privacy challenges (e.g., [46, 62]), challenges that result from shifting motivations at different life stages [20, 111], and challenges due to differences between user expectations and platform functionality [32].

In HCI, extensive attention has been given to *digital legacies*, the collection of materials that carry values and meaning that are passed down and/or otherwise continue to represent the deceased after their death. A recent literature review of the field by Doyle and Brubaker [25] identified four foci of this scholarship: digital identity (i.e., how legacies can continue to represent the deceased in intentional and unintentional ways), engagement with digital legacies (i.e., studies on use and user perceptions), putting to rest (i.e., concerned with issues around preservation, disposition, and disposal), and the integration of technology into existing legacy practices [25].

In most cases, it is the digital traces people leave behind that give rise to legacy crafting and memorialization opportunities. The sources of legacy content vary – from burner accounts [39] to personal archives [53, 101] – but social media content (especially memorialized profiles) has received the most scholarly attention (e.g., [4, 12–14, 16, 32, 33, 35, 73, 76, 105, 106]).

The persistence of identities has benefits and challenges: Digital legacies can facilitate healthy grieving [15, 17, 101] and maintain connections to the deceased [14, 35, 41]. However, receiving a large and uncurated set of content can be overwhelming for loved ones [46], and may provide (for better or worse) an uncensored version of the deceased [42]. Loved ones may also find themselves overwhelmed as they become the contact point for the deceased's various online networks [13, 15].

Across many studies, participants describe uncertainty around the use, care, and understanding of digital legacies. Ambiguities often arise when data created for one purpose – e.g., personal archives [46, 53] or interpersonal communication [32, 33, 41, 72, 111] – are repurposed post-mortem. In contrast to traditional legacies where there are familiar norms and practices, Pfister [91] argues that common norms around digital legacy have not yet emerged, an issue that AI will certainly exacerbate.

2.2 AI Afterlives

A handful of tech-savvy individuals have attempted to create interactive memorializations of loved ones by ingesting their digital (or digitized) content such as emails, journals, videos, photographs, or other autobiographical media. In her graphic novel *Artificial: A Love Story*, Amy Kurzweil [57] documents how her father, futurist Ray Kurzweil, created a chatbot to embody the memory of his deceased father, Fred Kurzweil, in a form he dubs "Fredbot" [115]. Fredbot interactively responds to questions from his descendants, but only by sharing exact quotes from the materials Fred left behind such as letters digitized by his family. In another well-publicized incident, the engineer Eugenia Kuyda created an app to preserve the memory of her best friend, Roman, who died unexpectedly in an accident, training a neural network on text messages her friend had sent her to create a bot named, suitably, "Roman." Unlike Fredbot, which was available only to immediate family, the Roman bot was made available on social media and app stores for public interaction, resulting in mixed reactions from friends and family of the deceased [81]. While Fredbot and Roman bot represented private citizens, AI has also been used to re-create public figures in various ways. For instance, in November 2023, the surviving members of The Beatles released a new song, "Now and Then," [7] using AI to enable the deceased John Lennon to sing along with his living bandmates [97]. The musician Laurie Anderson collaborated with the University of Adelaide's Australian Institute for Machine Learning to create a chatbot based on her longtime