On Grief and Griefbots

Abstract: Griefbots are chatbots designed to assist individuals in coping with the loss of a loved one by offering a digital replica of the departed. Navigating grief is a deeply transformative and vulnerable journey intricately tied to one's well-being. Do griefbots aid in the grieving process, or do they unintentionally, or not, complicate it? To address these questions, this paper blends insights from philosophy and neuroscience to explore the nature of grief as a means to clarify the ethical dimensions surrounding the use of griefbots.

Large Language Models are all the hype right now. Amongst the things we can use them for, is the creation of digital personas, known as 'griefbots', that imitate the way people who passed away spoke and wrote. This can be achieved by inputting a person's data, including their written works, blog posts, social media content, photos, videos, and more, into a Large Language Model such as ChatGPT. Unlike deepfakes, griefbots are dynamic digital entities that continuously learn and adapt. They can process new information, provide responses to questions, offer guidance, and even engage in discussions on current events or personal topics, all while echoing the unique voice and language patterns of the individuals they mimic.

Numerous startups are already anticipating the growing demand for digital personas. Replika is one of the first companies to offer griefbots, although now they focus on providing more general AI companions, "always there to listen and talk, always on your side". HereAfter AI offers the opportunity to capture one's life story by engaging in dialogue with either a chatbot or a human biographer. This data is then harnessed and compiled with other data points to construct a lifelike replica of oneself that can then be offered to loved ones "for the holidays, Mother's Day, Father's Day, birthdays, retirements, and more." Also, You, Only Virtual, is "pioneering advanced digital communications so that we Never Have to Say Goodbye to those we love."

The list of companies with suggestive names and not-so-modest mottos could go on. But almost all of them aim to help people deal with the passing away of their loved ones by providing a means to keep lost relationships alive. When we zoom out, it becomes evident that the 'grief tech' sector subscribes to the same *laissez-faire* ethos in its approach to technology

development and deployment as the rest of the digital tech industry. Developers often embrace a 'solve as we go' perspective when it comes to the impact of their products on users. But grief can be a transformative and vulnerable experience, which profoundly shapes people's well-being. For this reason, we need a nuanced exploration of griefbots. Before doing this, though, we should have a better grasp of grief, its nature, and its purpose. Is grief something we should do away with or is there some value in it? What can philosophy tell us about grieving?

A brief philosophical history of grief

The constellation of problems, questions, and problematizations death gave rise to throughout the history of philosophy is impressive. Despite this richness, one particular death-related event has remained largely ignored and undertheorized: grief, or the experience of the death of those we love. The marginalization of grief could be a legacy of ancient Greek and Roman philosophers. Even though grief gave rise to a new literary genre, consolation letters, the prevailing impression when reading them is that we should either avoid grieving altogether or minimize it as much as possible. Take, for example, Socrates, who scolds his friends, in *Phaedo* (although not a consolation letter *per se*, this dialogue can function as one), when they start crying at the thought of his imminent death: "What a way to behave, my strange friends! Why, it was mainly for this reason that I sent the women away, so that they shouldn't make this sort of trouble". Grief is a weakness. Epictetus advises in his *Manual*: "When giving your wife or child a kiss, repeat to yourself, 'I am kissing a mortal.' Then you won't be so distraught if they are taken from you." Grief is redundant. Seneca, in one of his letters to Lucilius, is a bit more understanding in claiming that: "Let not the eyes be dry when we have lost a friend, nor let them overflow. We may weep, but we must not wail." Grief should be temperate. All schools of ancient philosophy, despite their differences on almost all other matters, seem to converge on the fact that grieving is something we should minimize if not entirely eradicate.

The ancients' *mefiance* of grieving is understandable, after all, it *feels* bad. Most people experience sadness, a feeling of longing for the other person, some feel loneliness, guilt, or anger when someone close dies. Others, still, feel disconnected from their own life and the world as a whole. C.S. Lewis, in *A Grief Observed*, describes his grief for his wife as "an invisible blanket

between the world and me". Grief is often associated with emotions that take a toll on overall well-being, just like challenging mental conditions that make it hard to function. It is not a state we want to go through, nor is it something we would wish for our loved ones after we die. Hence, advice on how to minimize or eliminate grief seems more than welcome.

Grieving as learning

Over time, there has been a shift as both scientists and philosophers have sought to understand the nature of grief. Interestingly, their research converged around a set of fundamental questions: What if we were to embrace grief? Could it be that grief, rather than an insurmountable burden, is a form of learning that equips us to confront the void left in our lives by the departure of those we hold dear?

Our brain is a prediction machine that maintains relatively stable representations of both internal and external environments. This is why after someone we love dies, we still have the urge to give them a call or go talk to them. Forming a deep bond with someone will encode in our brain the expectation that they'll always be around. At the neurocognitive level, grieving is the brain's effort to update its model of the world. It arises from the conflict between two streams of information: semantic knowledge, representing the persistence of the attachment figure's existence, and episodic knowledge of the death of the person, such as memory of the funeral or of the moment we got the news someone we loved died. The 'gone-but-also-everlasting model' advanced by the clinical psychologist Mary-Frances O'Connor reveals the learning function of grieving. Our brains need time to replace the old habits associated with the person who died and acquire new automatic behaviors. Similarly, it takes time for our brain to develop a new perspective on ourselves, the world, and our future.

Michael Cholbi's recent book, *Grief. A Philosophical Guide*, mirrors these insights, portraying grief as a transformative experience that can help us reconstruct life's meaning in the face of death. We grieve for those in whom we are eudaimonically invested, whose existence is incorporated into our practical identities. We grieve for those we love, and loving means that our well-being is shared with others. This is love's bond, as Nozick put it in *The Examined Life*: in every loving relationship, there is a *me*, a *you*, and a *we*. When someone we love dies, we are left

with just a *me* that used to be defined by the *you* and the *we*, and this new situation does not make too much sense. Death forces us to confront the question of how those we loved mattered to us, what we valued in our relationship, and what we will be missing. The retrospective dimension of grief offers a way to reconstruct our past, which can help us better understand ourselves. But this type of existential reflection generates a 'crisis' in our identities, one that makes us rethink ourselves. In its forward-looking dimension, grief forces us to reflect on how our lives will continue. Here lies grief's paradoxical nature: despite being a painful experience, it is valuable because it involves a reconstruction of one's image of oneself. It is a way of learning who we are, a form of self-knowledge.

Not every grief experience will wane with time. Some people do not want to overcome the sadness and despair provoked by the death of a close one, while others simply cannot do it. But if we look around, people almost always seem to be able to resume their lives and have meaningful relations after a grief episode. This is because, if we take psychologists and philosophers seriously, grief is learning. Learning to let go.

Griefbots reconsidered

Griefbots are poised to reshape the way we remember those we've lost. When we revisit videos, photos, emails, or letters of our departed loved ones, we typically encounter passive representations of the past that don't offer any signs of life. In contrast, griefbots are incredibly lifelike and present. They can respond to questions, comment on current events, make relevant jokes, and offer life advice based on the information we provide. They also can respond spontaneously and appropriately to users' emotions, constantly evolving and learning not just from the data they're given but also from previous interactions. The distinctiveness of griefbots lies in their unpredictability, which can make our interactions with them feel remarkably real. They're not just passive tools for remembering; they're a means of maintaining an active connection with those who are not here anymore and even creating new memories in the process.

There is, thus, a significant likelihood that griefbots could disrupt the grieving process. Because of their interactive nature, griefbots can create the illusion of an ongoing connection with the deceased. This can make it more challenging to come to terms with and accept the

reality of a loved one's passing. It's not necessarily that people will eventually see those who died as continuing to exist in a digital realm. But these technologies could give rise to cognitive dissonance: on the one hand, we know the person we loved died; on the other, we can feel as if they are there, precisely when talking with a griefbot that can respond in like manner. This cognitive dissonance can not only affect well-being but also stop people from learning to live meaningful lives in a world without a loved person. Griefbots have the potential to blur the line between living in an alternate reality where our loved ones still exist somehow and living in the present moment while remembering the reality of their absence as well as why they mattered to us. There are also more tangible worries, mainly having to do with how companies' financial incentives might shape the development of these technologies and how they might impact users. For example, can griefbots be used for advertising? Will they be used for manipulation?

But there's a silver lining to every cloud, even in the case of griefbots. To uncover it, we need to don the lenses of ancient philosophers and view death and grief through their perspective. At first glance, it might seem that Epictetus and Seneca are suggesting that we shouldn't make other people central to our lives and well-being due to their impermanence. This is one way to look at it. A more charitable interpretation is that there's no harm in forming deep connections with others and making them central to our lives, as long as we constantly remind ourselves of the inevitability of their eventual death. In doing so, we can concentrate on what truly matters, without getting sidetracked by everyday trivialities. This opens the possibility of wholeheartedly investing in our relationships without being crushed by the weight of their loss. By keeping alive, through our actions and relationships, the qualities we valued in those we've lost, by keeping alive what we were in relation to those who passed away, we can express our love and gratitude for them without succumbing to despair.

But memory is fickle and the danger lies in losing how those who passed away actually were. And in doing so, we also lose what we were in connection to them. This is because those who matter to us contribute enormously to how we construct and understand ourselves. Photographs, videos, emails, or voice messages remind us of the image of the person who died. They cannot recreate the *we* that the love's bond creates. But griefbots, due to their interactivity, can help us not only remember the other person but also to recreate what we were in connection with them. Just envision the possibility of occasionally having a brief chat with a chatbot that flawlessly emulates your father, mother, or anyone whom you loved and who has passed away.

This experience could function as a means not to preserve their existence, but rather to keep alive that part of yourself that was molded by your relationship with them and which can fade with time. The effects of conversations with a griefbot can extend beyond personal satisfaction; they can also have moral significance. Through this interaction, we have the opportunity to revisit and rekindle what we treasured in the people who died and the connections we shared with them. For example, by engaging with a digital entity that replicates my father's voice, I can be reminded how good it felt to be protected, and how much comfort his calm and control offered me. And if these are the things that I love and miss, maybe this is what I should also build in my relationships with others. However, it's important to approach dialogues with griefbots with a sense of separation, rather than expecting a complete reconnection. The goal is not to replace the profound emotional depth of a living person, but rather to use griefbots as a tool for re-learning what we love.

The impact of griefbots, whether they hinder acceptance of a loved one's passing or provide comfort and ease grieving, depends not only on individual responses but also on the way these technologies are designed. Grief, which is a time of vulnerability, can be an opportunity for tech companies to exploit these emotions for their benefit, sometimes at our expense. And they have a bad track record of that. It's time we adopt a proactive approach to address ethical concerns in technology design, one that anticipates potential harms these technologies might cause and seeks to mitigate them before they reach the market. This holds particularly true for artifacts that deal with life and death. The values embedded in technology design should prioritize transparency in how griefbots operate, including how they use the data we feed them with, and constant reminders that they lack intention and consciousness. This way, we can ensure that griefbots serve their intended purpose without inadvertently causing harm.