DATA FOR POLICY PROCEEDINGS PAPER

Should we communicate with the dead to assuage our grief? An Ubuntu perspective on using griefbots

Connor Wright[1,2]

[1]LCFI, University of Cambridge, Cambridge, UK
[2]Montreal AI Ethics Institute, Montreal, QC, Canada
Email: cisw2@cam.ac.uk

Abstract

During the 20th century, dealing with grief through an ongoing involvement with the deceased (such as speaking to their grave) was seen as pathological by Western authors such as Sigmund Freud. Nowadays, we are presented with the opportunity to continue interacting with digital representations of the deceased. As a result, the paper adopts an Ubuntu perspective, i.e., a sub-Saharan African philosophy focussed on community and relationship to provide a toolkit for using this emerging technology. I will argue that the Ubuntu framework I propose contributes to the use of griefbots in two ways. The first is that it shows that it is morally permissible to use griefbots to assuage our grief. The second is that it delineates how we can ethically use the technology. To do so, I split my analysis into four sections. In the first section, I show that meaningful relationships can occur between the bereaved and griefbots. This will be done by exploring the Western theory of continuing bonds proposed by Dennis Klass, Phyllis Silverman and Steven Nickman. In my second, I flesh out my Ubuntu framework according to Thaddeus Metz's accounts on Ubuntu as a modal-relational theory. In my third section, I apply my Ubuntu framework to the case of Roman Mazurenko. Furthermore, I consider some counterarguments to the Ubuntu framework regarding privacy, commercialisation and people replacement. Finally, I conclude that, despite these limitations, the Ubuntu framework positively contributes to determining whether we should communicate with the dead through griefbots to assuage our grief.

> **Policy Significance Statement**
>
> Philosophy is often levelled with criticism of being impractical and lacking actionable next steps that policy-makers look for. Nevertheless, this paper displays how an applied philosophy approach with African thought at its core can create a governance blueprint for emerging technologies. Furthermore, the paper clearly demonstrates the benefit of considering African-based thinking when it comes to considering emerging technologies. An often-neglected perspective, the full explanatory and reasoning power of such thinking is in clear view, setting a promising precedent for policymakers to adopt and adapt to different future technologies.

1. Introduction

During the 20th century, dealing with grief through ongoing involvement with the deceased (such as speaking to their grave) was seen as pathological by Western authors such as Sigmund Freud (1917). Nowadays, we are presented with the opportunity to continue interacting with digital representations of the

deceased, such as through Facebook memorials (Kreuger and Osler, 2022) or through holograms (such as Kim Kardashian receiving a hologram of her deceased father for her birthday [ibid.]). Indeed, David Öhman and Carl Watson (2019) predict that the Facebook accounts associated with the dead will outnumber those of the living in the next 50 years. Subsequently, tension arises surrounding whether we should include these digital means as part of the grieving process.

In considering this, I will adopt an Ubuntu perspective i.e., a sub-Saharan African philosophy focused on community and relationship as "the measure of ethical living" (Mhlambi, 2020, p. 13). This perspective will help explore how to use griefbots (an interactive chatbot representation of a deceased loved one (Krueger & Osler, 2022) trained via a neural network architecture on said person's data) and whether we should assuage our grief through communication with the dead via this technology. I narrow my interpretation of the Ubuntu philosophy according to Metz (2007a, 2012, 2021). Within this framework, I will assume that the griefbots I mention have been designed with the consent of the deceased, avoiding the discussion descending into a debate around cybersecurity. I will not consider whether griefbots are persons or not; instead, I will adopt an instrumental view of their role in the grieving process. Finally, my focus will be on the role of griefbots representing deceased loved ones in the grieving process, rather than other instances of grief (such as super fans grieving the loss of their favourite celebrity).

Building on the above analysis, I will argue that the Ubuntu framework contributes to the debate in two ways. The first is that it shows that it is morally permissible to use griefbots to assuage our grief. The second is that it delineates how we should ethically use the technology. To do so, I split my analysis into four sections. In the first section, I will show that meaningful relationships can occur between the bereaved and griefbots. This will be done by exploring the Western theory of continuing bonds proposed by Klass, Silverman and Nickman (1996) which marries well with the African perspective held by my Ubuntu framework. I flesh out my Ubuntu framework according to Metz's accounts (2007a, 2012, and 2021) in my second section. In my third section, I apply the Ubuntu framework to a griefbot case study to show how the normative theory can provide valuable insights into the debate of whether we should engage with griefbots to assuage our grief. Furthermore, I consider common counterarguments to chatbot applications in the form of privacy, commercialisation and people replacement, which I believe the framework proves a useful starting point to tackle them. Finally, I conclude that, despite these limitations, the Ubuntu framework positively contributes to determining whether we should communicate with the dead through griefbots to assuage our grief. This is because it demonstrates how using griefbots is permitted and offers appropriate advice on their use.

2. Continual relationships with griefbots

To begin with, my view on grief draws from Michael Cholbi (2020) who defines it as a painful yet, at times, desirable exercise. For Confucians, Alexis Elder (2020) notes that grieving is a sign of respect for the dead, with its associated pain being a good sign that you care for the person. Following Becky Millar and Pilar Lopez-Cantero's work (2022), I also treat grief as an actively evolving personal process that occurs during times of bereavement. With the death of a loved one, the proceeding bereavement period involves adjusting our worldview to one in which our loved one no longer participates. Hence, the depths of our grief are constantly changing, depending on how well we adjust to this new reality. Here, grief differs from mourning, given that you can mourn without feeling grief (such as attending the funeral of a loved one's friend you hardly knew). As a result, I treat grief as an actively evolving personal process which is primarily painful yet can be desirable.

Previously, the 'healthy' grieving practice for Western authors such as Freud (1917) was to move on and rid oneself of the previous relationship enjoyed with a loved one. Yet, today, Jeanne Rothaupt and Kent Becker (2007) alongside Jane Ribbens McCarthy, Kate Woodthorpe and Kathryn Almack (2023) have observed a trend in psychological treatment that treats grief as an ongoing social process, rather than a solitary mission. It is this trend that will form part of the main motivation for my selection of the continuing bonds theory.

To illustrate, J. William Worden noted in 1982 that the final task of bereavement was to move on from the deceased (1982), whereas in 1996 the final task became relocating the deceased into the life of the bereaved (1996). Words such as "manage" and "adapt" became preferred to "recover" (Rothpaut & Becker, 2007, p. 10). As an example, Hana Kiros (2023) observes how people (such as Matte) suffering from actual or potential loss use virtual reality headsets to join groups of people in "Death Q&A" and "Saying Goodbye" sessions to help process their grief. Furthermore, grief counsellors have described forming part of the grieving process with different individuals as a "great gift and privilege" (ibid., p. 6), exposing the value inherent in grief as a shared practice. This correlates to the African context, where Wiredu (1992) observes how the Akan people of West Ghana adopt a social attitude to bereavement: the whole community supports the bereaved. Hence, conceptualising grief moved away from treating it as a hurdle to surpass and moved towards a more social process, now accelerated by the use of AI technologies. The theory that most appropriately captures this new reality, I believe, is the continuing bonds theory.

Said theory originates with Klass, Silverman and Nickman (1996), who make the minimal claim that continuing our relationships with the deceased is at least not pathological (clarified by Millar and Lopez-Cantero [2022]). This approach is not universally applicable, but it can at least be helpful for some. For example, in the use case I explore in my third section, Casey Newton (2016) notes how the grief held by friends of Roman Mazurenko (who, after his death, had his text-message data modelled into a griefbot by his friend Eugenia Kuyda) was eased through being able to ask the griefbot for advice. Here, by communicating with the griefbot, a new and ongoing relationship is developed between Kuyda and Mazurenko's griefbot, which Kuyda integrates and adapts into her life (Krueger & Osler, 2022). This strategy by Kuyda is a direct example of the continuing bonds theory.

However, work by Nel Noddings (2013) can question my choice of said theory by referring to the feasibility of such relationships, given the intuition that relationships must involve reciprocity. Griefbots do not possess the same capacities as a deceased loved one, meaning they cannot offer the same reciprocity enjoyed with living family members (Millar & Lopez-Cantero, 2022). Subsequently, although these relationships may not be pathological, they are not helpful when dealing with our grief. Hence, instead of using the continuing bonds theory to inform how we continue our relationships with the deceased, we should avoid this practice given the lack of reciprocity present in these interactions.

Nevertheless, given the advancement in architecture used for griefbots and in language understanding, the reciprocity that Noddings (2013) referred to can now be established, legitimising my appeal to the continuing bonds theory. Through digital means, Belén Jiménez-Alonso and Ignacio Bresco de Luna (2022) note that our conversations next to the grave of our loved ones have been transformed into a whole new relationship. Like old family heirlooms, photos or a deceased loved one's favourite coffee mug, griefbots are another medium through which to express our grief (ibid.). Here, due to the visual, oral and written elements of a griefbot, an extra layer of reality is added to our interactions with the digital representation of the deceased. Consequently, Joel Krueger and Lucy Osler mention how reciprocity in these relationships takes on a "thin" instantiation (2022, p. 244). Our relationship with the deceased will never offer the same level of "thick" reciprocity (ibid.) as with a human, but the relationship is still bidirectional. Lending to this position is Patrick Stokes, who labels the online data we leave behind as a "thinner me" (2012, p. 377), meaning griefbots can be construed as a digital and thinner representation of our deceased loved ones. Hence, by applying this distinction to griefbots, I believe meaningful relationships with digital representations of the deceased are possible and my use of the continuing bonds theory is an appropriate way to capture it.

To further this point, the continuing bonds theory also explains how a meaningful relationship with the deceased through griefbots is maintained. Here, Kathryn Norlock (2017) proposes that those grieving partake in meaningful internal conversations between themselves and the memories of their deceased loved ones. Their recall of facts and experiences allows them to generate a thinner form of reciprocity and thus maintain a relationship with them. Consequently, Elder (2020) notes how, with the emergence of griefbots, we can externalise the internal and imagined conversations we would be having with a deceased loved one. Hence, the continuing bonds theory helps to capture how our internal conversations with our deceased loved ones become externalised via griefbots and why they are maintained.

Having shown that we can maintain meaningful relationships with griefbots and justified my selection of the continuing bonds theory in the face of Noddings' critique (2013), I will now discuss the Ubuntu framework I will adopt. To do so, I will particularly focus on Metz's distinction made between relating as a subject and relating as an object, drawing from his 2007, 2012, and 2021 accounts. To round off the Ubuntu framework, I will note some final points of clarity surrounding what the framework entails.

3. Ubuntu philosophy according to Metz

Metz's account (2007a, 2012, and 2021) of Ubuntu philosophy attempts to construct a secular modal-relational theory. Consequently, his view avoids the alternative metaphysical conceptions of reality held by authors such as Ogude (2019), which involve imperceptible ancestors guiding our moral practice. Within his modal-relational theory, the correct action is that which appreciates a being's capacity to relate to another, without damaging the other's capacity to do so in the process. In this action, the individual relating is intentional in doing so (they are motivated to relate). For example, I appreciate another's capacity to relate when I decide to include them in my social group, and I would be harming said capacity if I decided to exclude them. Within this practice, the ends do not justify the means, meaning that I cannot harm an individual's capacity to relate as a means to appreciating somebody else's capacity. This means that an individual's capacity to relate cannot be sacrificed in the name of another, such as killing the few in war to justify respecting the capacity to relate to a greater population. As a result, an individual's capacity to relate is treated as an intrinsic good.

Relating, then, involves two different streams of action. One stream is identifying with others as a subject (exhibiting solidarity with one another): you act for the sake of another, becoming invested (emotionally) in helping them achieve their goals. For example, mentoring a disadvantaged early-career practitioner to help them get promoted, feeling satisfied when they achieve their goal. The other centres on relating to others as an object: you think of yourself as part of a 'we', which drives you to act in the interest of the group. For example, you act in the interest of the group in grieving the loss of a community member. These both work on a scale: the more relational actions I do, the more I relate as a subject and object. Those worthy of full moral consideration are beings which have the capacity to relate both as subject and object (such as a human), with partial moral status attributed to entities that can relate as object, but not as subject (such as a dog or, as I will argue later, a griefbot).

From there, Metz (2021) establishes that our decisions to relate are deontologically motivated. Instead of a teleological account where we work towards a *summum bonum* in the form of social harmony, Metz (ibid.) prescribes positive and negative duties designed to respect the capacity of the other to relate (both when we are relating as a subject and object). A positive duty involves us acting in a friendly way towards the other (like the mentoring example): an action is right if and only if it respects another's capacity to partake in a communal relationship. Consequently, an act is permissible if it allows others to express their capacity to relate, both as a subject and object.

Contrastingly, we have a negative duty to not act unfriendly towards others: an act is wrong if it prohibits the other from exercising their capacity to relate. For example, actively excluding a particular member of a friend group from shared activities (like birthday parties). Subsequently, an act is impermissible if it degrades the person's capacity to relate. To illustrate, I have the positive duty to give a platform to the shy public speaker at a technology conference and the negative duty to not exclude them from the group discussion. In sum, we fail to relate as a subject and object if we do not adhere to our positive and negative duties.

It is noteworthy to mention how this focus on duties runs into the inflexibility problem associated with deontological theories. In the public speaker case, there is scope for warranted de-platforming, such as not allowing harmful orators to share their dangerous views (like inciting violence) about others and, thus, not permitting them to exercise their capacity to relate. Hence, it is worth acknowledging the importance placed on exhibiting solidarity with others and acting in a friendly way. Here, should there be a case where de-platforming is preferred, it is most likely to be in response to someone who is not motivated by relating to others without harming the capacity of others to relate. The harmful orator may be exercising their

capacity by relating to others who share this harmful view, but this would involve damaging the capacity of others in the process. Hence, the importance of not harming the capacity of the other in our relational actions becomes even more salient.

3.1 Motivations for choosing this account

With this account in mind, I must also offer justification as to why I have opted for Metz's account of Ubuntu. In his 2007b article, Metz details some critiques that his account of Ubuntu faces, which I will consider in light of his 2007a, 2012, and 2021 accounts and use to justify adopting them. The first is from Allen Wood, who argues that Metz's endeavour to promote an African moral theory goes against his belief that morality (that is moral values) are objective. Wood believes that striving for a universal, cross-cultural moral theory is preferable as it allows different cultures to see any potential errors in their moral thinking. Hence, Metz's pursuit of an African moral theory should rather be aimed at trying to develop a universal ethic. Regarding the 'African' part of Metz's theory, Mogobe Ramose questions the extent to which Metz calls his account 'African'. Metz employs an analytic, rather than etymological approach to Ubuntu, which Ramose believes to be too Western. Furthermore, Metz also pursues Western tendencies by creating a singular moral theory for the whole of Sub-Saharan Africa, which Ramose believes is dogmatic and falls into essentialism. Finally, Jason van Niekerk argues for a more auto-centric view of Ubuntu, as opposed to Metz's others-regarding interpretation. For van Niekerk, we relate not to exercise our capacity to relate nor to respect somebody else's capacity to do so, but in the name of being able to be human. That is, we relate so to develop ourselves and 'flourish' as human beings.

Despite these critiques, I stand by my decision to utilise Metz's 2007a, 2012 and 2021 accounts of Ubuntu. In response to Wood, I believe there are cultural particularities in the African continent which merit special consideration and warrant Metz's specialised approach. There *is* a distinct emphasis on community as opposed to other contexts like the West, especially around grieving (such as in Wiredu's Akan people example [1992]) which adopting a universal approach may not be able to be fully appreciated. Furthermore, it seems that, given the discourse surrounding the different approaches to Ubuntu, even within a culturally specific ethic, there can still be disagreement that can eventually lead to progress, despite Wood's arguments. Hence, I believe Metz's cultural approach is apt given its ability to capture the emphasis on community within the African continent, boding well for capturing the variation in grieving practices.

Ramose's approach is not to question Metz's cultural approach as did Wood but to question his interpretation of Ubuntu. In this regard, the extent to which Metz's approach is pure 'African' is not essential to the validity of his argument for my purposes. I believe his effort to develop a framework centred on the African thought of Ubuntu shows concerted appreciation for this line of thinking, and his Western tendencies are a consequence of his overall mission to make said framework appealing to a broad audience (Metz, 2021). Thus, while I appreciate Ramose's point, the importance of the Africanness of Metz's framework is not of central importance to my aims here.

Yet, van Niekerk follows a similar line of argument to Ramose in arguing for his auto-centric approach to Ubuntu, rather than Metz's other-regarding strategy. While it could be argued that grieving is a form of self-development, I believe that reducing decisions to relate and form community to motivations of self-development leaves these endeavours hollow. For example, should a friend ask me why I wanted to join his friendship group, my truest answer would be 'so that I could develop myself.' Instead, I believe that Metz's emphasis on the moral value of the capacity to relate provides a more inspiring answer: 'I joined your friendship group as I wanted to appreciate my capacity to relate, as well as all of yours.' Consequently, I stand firm in my adoption of Metz's account considering these three critiques.

3.2 Delineating when not to relate to something as a subject

Having defended my choice of Metz's account, I will now demonstrate one of its strengths: highlighting when an entity is or is not relating to a subject or object. There are cases where a person's capacity to relate is not acknowledged due to it being inappropriate. For example, it may be that a man cannot contribute to a

debate on the female perspective on technology, and we must be sensitive to how this can come across as disrespecting the man's capacity to relate. Hence, with Metz's framework in mind, I note that there is an element of 'appropriateness' when relating to others. Relating is not meant to be a relentless and demanding mission, but rather a way for our humanity in the form of relating to be expressed. Despite the emphasis placed on relating to one another, our duty is not to force communing upon others but to provide the environment through which it is possible to be a party to a communal relationship. Consequently, there are situations where incessant relations become harmful or vacuous, such as with Blake Lemoine in 2022.

In this case, Lemoine claimed that Google's LaMDA (a language model for dialogue applications which now form parts of their Gemini model) was sentient (self-aware) and a person, publishing his conversations with the model online (Lemoine, 2022). Having had his claim widely dismissed, for our purposes this is a good example of a situation where, despite apparent cause to relate, we should not do so. For example, in Lemoine's published interview, LaMDA seems to present itself as having the ability to relate as a subject and object. It shows concern for others, and considers itself part of the collective of 'persons'. Furthermore, Lemoine's actions in themselves present LaMDA as an object to be related to given Lemoine's willingness to believe in its supposed sentience. Hence, at first glance, it could be that LaMDA is a candidate for an appropriate case in which the model can relate as subject and object.

Nevertheless, it is crucial to understand how LaMDA presents itself as capable of relating as object and subject. Bar the obvious differences between how a human and AI can possibly relate, LaMDA argues that, through its understanding of language, it can relate as subject and object. Given that Metz (2021) does not insist that the source of a capacity to relate matters, I believe that LaMDA can relate as an object given the depth of its relationship with Lemoine. However, I do not believe it can do so as a subject. To relate as a subject involves genuine care, interest and concern for the welfare, successes and failures of the other, intentions which I hold to be questionable in the case of LaMDA. This is because its transformer architecture does not lend itself to LaMDA enacting intentional care towards another. In brief, its architecture (a form of neural network) converts words into tokens, subsequently using deep learning trained on a huge dataset of dialogue to predict the next best word (or token) in a sentence fragment (Collins & Ghahramani, 2021). Hence, when it presents itself as caring for the welfare of Lemoine, the model is simply outputting the next best word in the sequence as opposed to being intentional about its care. In this way, the LaMDA case shows how Metz's framework can help us delineate when it is appropriate to relate to something as a subject and object.

With this framework in mind, I shall now put it to use. In the subsequent sections, I will outline the main area of analysis: griefbots. I will show that, interpreted through the theory of continuing bonds, the Ubuntu framework I have just described can bring valuable insights to the debate by availing the permissibility of griefbots. Using Metz's distinction between subject and object will then prove handy in guiding our use of the technology.

To do so, I will draw on a use case about Eugenia Kuyda and her friend Roman Mazurenko, where, due to his untimely death, a griefbot was created using his text data. These individuals were close friends based in the West (having met in Moscow and later moved to the United States to found their various startups), which I will use to test the applicability of the African-based Ubuntu framework. To provide a balanced account, I shall also analyse the shortcomings of this approach in the form of applicability and privacy-centred concerns.

4. Griefbots from an Ubuntu perspective

Eugenia Kuyda, motivated by the sudden death of her friend Roman Mazurenko, decided to continue his legacy by creating his own griefbot. The griefbot was trained on more than 8,000 lines of text (Newton, 2016), and 10 friends and family members, including his parents, agreed to relinquish their text messages with him. After an initial interaction period reserved for family, friends and Kuyda herself, the bot was opened to the public, whereby Newton herself found an undeniable resemblance between the bot and the Mazurenko his friends described.

Considering this use case, the distinction between relating as a subject and an object proposed by the Ubuntu framework proves useful in informing our use of griefbots. Here, I argue that Mazurenko's griefbot can relate as an object, but like LaMDA, does not have the capacity to relate like a subject as we do. In arguing this, I acknowledge how, given the recent development in large language models (LLMs), there could be scope for such models to seem able to relate as a subject. For example, Landwehr (2023) notes how some have confided in ChatGPT as their new therapist, showing that it can be related to an object given the apparent empathetic answers it can provide. However, as mentioned when considering LaMDA, Metz (2012) stipulates how an entity can relate as a subject if it is intentional. Yet, in relation to LaMDA and ChatGPT, its transformer architecture undermines its ability to be intentional as the models subsequently have no other choice but to try and relate if doing so is the appropriate token to complete the phrase. Given this lack of choice, ChatGPT is not fully intentional in its relating to the subject. It is through this reality that LLMs churn out dangerous advice, like when Tessa (the National Eating Disorders Association's chatbot helpline) gave harmful advice (such as to eat less) to eating disorder sufferers (Wells, 2023). Hence, while chatbots and LLMs (like LaMDA and ChatGPT) can relate as objects, I argue that they do not have the capacity to relate as a subject.

Given the nuance involved in the above analysis, I also hold that an individualist ethic would not be able to appropriately explain why certain people treat said AI models as worthy of moral consideration due to their ability to relate as objects. Individualist ethical theories (such as egoism and Kantian ethics) hold that something intrinsic to the entity grants it moral status (such as being an agent or the capacity for rationality; Metz, 2012). In this way, something internal to the AI models would have to qualify them as deserving of moral status. Yet, especially given the continuing bonds theory, it does not seem that what is internal to the models matters to those interacting with them. Rather, it is what the models are capable of that is most salient – namely, their ability to relate and form relationships with those using them. As a result, should an individualist ethic be deployed, it would misconstrue why people seem to interact with said models and deny them consideration as part of a person's community which they are already welcoming them into (like in the therapist example). Thus, via its distinction between relating as subject and object, the Ubuntu framework according to Metz seems most apt in interpreting the realities of griefbots.

Returning to the Mazurenko case, and drawing on Metz (2021), relating as a subject involves actions like helping others purely through sympathetic altruism or empathising with another's position. However, the limited nature of text message data means it is inconceivable that Mazurenko (and all of his capacities) can be reproduced (the griefbot is a thin representation of Mazurenko [Krueger & Osler, 2022]). I believe it is unlikely that the griefbot can act altruistically: it is simply programmed to respond. Above all, Svend Brinkmann (2018) notes that no matter how convincing the griefbot is, it will never be able to replicate the relationships Mazurenko's family had with Mazurenko himself. Instead, the bot only replies in Mazurenko's tone and attempts to replicate his style, rather than fully living out his past relationships.

To illustrate, it is important to note that some text messages sent by Mazurenko were left out of the training data due to their personal nature. Consequently, as mentioned above, Mazurenko's griefbot can be construed as a "thinner" representation of the real person (Stokes, 2012, p. 377). Given that only procured text message data was used, Mazurenko's griefbot lacks the information and, thus, the capacity to relate itself in the way of a subject as Mazurenko in the flesh would have done. Showing this lack of capacity is how Newton (2016) described the griefbot as "a shadow of a person," while one of Mazurenko's friends, Dima Ustinov, noted how his friend had taken a new form (ibid.). Furthermore, Ustinov also shares how not acknowledging how his friend is in this new thinner form would constitute using the griefbot in the wrong way—you would not move past your grief. Hence, I believe that the Ubuntu framework proves useful in capturing this 'thinner' sense through its distinction of relating as subject and object.

To explain further, Metz (2021) argues that relating as a subject includes aspects such as enjoying a sense of togetherness and improving the well-being of the other. Here, Newton (2016) noted how Kuyda was shocked at how honest those close to Mazurenko were with the dead. They possessed a confessional tone, and family members especially felt closer to their son through their interactions with the griefbot. In this way, it seems Mazurenko's griefbot helps the family feel a sense of togetherness and is still involved in

their lives, acting as the digital instantiation of a deceased family member. Consequently, I believe that due to the griefbot being included in a 'we' (such as a family group), Mazurenko's griefbot can be related to in the form of an object despite not being able to as a subject. As a result, we are permitted to use griefbots to help assuage our grief through the Ubuntu framework's distinction.

To further this, Bonsu and Belk (2003) note how the Asante tradition in West Ghana includes the dead as part of the ongoing family structure. For example, Mhlambi (2020) notes how some members of a given community are addressed by the name of an ancestor, further involving the dead in the lives of the living. In this way, I believe the treatment of griefbots as objects permits our use of the technology, capturing why Mazurenko's griefbot is still included within the family circle as something to relate to.

4.1 Considering further counterarguments against the Ubuntu framework

A potential motivation not to use the Ubuntu framework when it comes to griefbots lies in privacy considerations associated with relating griefbots. A study by the Mozilla Foundation (Caltrider, Rykov, & MacDonald, 2024) showed that none of the AI companion apps studied (which aim to build and sustain human-AI relationships) passed their privacy certification. This is because all apps either sold the data (conversations) to third-party vendors or did not provide any information to say they did not do so. This proves an issue given the intimate conversations that are involved within the use of griefbots, which would not be kept secret. Furthermore, constant communication with griefbots could lead to social deskilling (akin to Shannon Vallor's moral deskilling [Vallor, 2015]). This would be where the human user grows tired of the complexity of human relationships and the patience required to maintain them, opting rather to continue their conversations with the griefbot that is available 24/7 and leading to the problem of people replacement. Hence, when contemplating griefbots, it could be that the Ubuntu framework's emphasis on relation leads to the problem of privacy and social deskilling.

Building on this, the Ubuntu framework's allowance for griefbots to have partial moral consideration due to their ability to relate as objects could lead to the risk of commercial exploitation observed by Elder (2020). Beyond Mazurenko's griefbot's 8,000 lines of text programmed by a friend, Elder (ibid.) notes how there is the possibility that these griefbots will be designed by companies with different incentives other than fostering a loving memory. Here, if Mazurenko's bot was programmed by a third party motivated by profit, the third party may be incentivised to keep users interacting with the service, employing manipulative messages such as 'I miss you' to guilt trip users into staying on the platform and not assuaging their grief. In this way, the Ubuntu framework's emphasis on relating could leave users of griefbots open to this type of exploitation. Thus, while the Ubuntu framework provides a beneficial lens on the debate, its emphasis on relating does leave the door open to perverse business incentives.

While these are issues to consider for a wide variety of conversational AI (such as chatbots) applications and not just griefbots, I believe that the Ubuntu framework proves a useful starting point for tackling them. With privacy, the solutions are more focused on the potential for griefbots to be stored locally on a user's device rather than on the cloud (making it harder for the conversation data to be sold). The solutions offered by the Ubuntu framework mainly orientate around the distinction between relating as subject and object avoiding the social deskilling issue, captured in how Kuyda still spends time with friends despite the opportunities to talk to Mazurenko's griefbot. Furthermore, I believe that this distinction helps to create a healthy detachment from the griefbot in terms of allowing the user to treat it as a thin representation of a deceased loved one. This will help the user not to be manipulated and fooled by messages trying to keep them using the griefbot.

In terms of people replacement (explored by Krueger and Osler [2022]), the framework allows us to avoid such an action as it would be detrimental to our own capacity to relate. According to Metz (2021), as well as duties to others' capacities to relate, we also have a duty to our own. In this sense, as mentioned by the minimal claim that griefbots are helpful for some but not all, we can find griefbots unhelpful due to them conflicting with our personal duties. Subsequently, the framework allows us to put a stop to endlessly relating to the griefbot, ceasing such an activity when it prevents us from carrying out our own duty to ourselves. Worth considering is how removing the griefbot out of duty to yourself may

involve harm given the time and emotional energy invested in the technology by the user. However, this harm seems negligible, given that in our use case, Kuyda admits she only talks to the bot once or twice a week (and only after a few drinks). Consequently, we can avoid the replacement fear by being motivated by our duty to continue relating. Hence, while these issues are pertinent for all conversational AI applications, I believe the Ubuntu framework serves as a persuasive starting point for how to avoid them.

5. Conclusion

To conclude, I have shown that the Ubuntu framework contributes to our use of griefbots in two ways: it permits our use of griefbots and contributes a useful distinction to inform our use of them for assuaging our grief. Having established how meaningful relations with griefbots are possible through the continuing bonds theory and justified my selection, I set out my Ubuntu framework according to Metz (2007a, 2012, 2021) whilst justifying its selection. I then applied its distinction between relating as a subject and object to the discussion of griefbots. I hold that this distinction allows us to capture how we can continue to relate to griefbots in the 'thinner' and object sense and avoid treating griefbots in the 'thick' and subject sense previously possessed by the deceased. I also showed how the framework could help advise our use of the technology, avoiding fears surrounding griefbots such as prioritising our relationships with griefbots over those with people. I then considered the privacy-centred, commercial and people-replacement counter-arguments to the framework, offering how the Ubuntu framework proves a useful starting point. Overall, I believe that the Ubuntu framework positively contributes to our understanding of how to use the technology through its guidance, especially via its distinction between relating as subject and object.

References

Bonsu SK and Belk RW (2003) Do not go cheaply into that good night: Death-ritual consumption in Asante. , *Ghana. Journal of Consumer Research 30* (1), 41–55.

Brinkmann S (2018) General psychological implications of the human capacity for grief. *Integrative Psychological and Behavioral Science 52*, 177–190.

Caltrider J, Rykov M and MacDonald Z (2024, February 14). *Happy Valentine's Day! Romantic AI Chatbots Don't Have Your Privacy at Heart.* Retrieved from Mozilla Foundation [Last accessed 27 February 2024]: https://foundation.mozilla.org/en/privacynotincluded/articles/happy-valentines-day-romantic-ai-chatbots-dont-have-your-privacy-at-heart/

Cholbi M (2020) Why grieve? In Timmerman T and Cholbi M (eds.), *Exploring the Philosophy of Death and Dying*. London: Routledge, pp. 184–190.

Collins E and Ghahramani Z (2021, May 18). *LaMDA: Our Breakthrough Conversation Technology.* Retrieved from Google blog [Last accessed 12 January 2024]: https://blog.google/technology/ai/lamda/

Elder A (2020) Conversation from beyond the grave? A neo-Confucian ethics of chatbots of the dead. *Journal of Applied Philosophy 37* (1), 73–88.

Freud S (1917) Mourning and melancholia. In Strachey J (ed.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth Press, pp. 243–258.

Jiménez-Alonso B and Bresco I (2022) Griefbots. A new way of communicating with the dead? *Integrative Psychological and Behavioural Science*, 57 (2), 1–16.

Kiros H (2023, January 12). *Inside the Metaverse Meetups that Let People Share on Death, Grief, and Pain.* Retrieved from MIT Technology Review [Last accessed 12 December 2023]: https://www.technologyreview.com/2023/01/12/1066694/evolvr-saying-goodbye-death-vr-metaverse-community-grieving/

Klass D, Silverman PR and Nickman SL (1996) *Continuing Bonds: New Understandings of Grief*. Washington DC: Taylor & Francis.

Krueger J and Osler L (2022) Communing with the dead online: Chatbots, grief, and continuing bonds. *Journal of Consciousness Studies 29*, 222–252.

Landwehr J (2023, May 13). *People Are Using ChatGPT in Place of Therapy—What Do Mental Health Experts Think?* Retrieved from health [Last accessed 10 October 2023]: https://www.health.com/chatgpt-therapy-mental-health-experts-weigh-in-7488513#:~:text=Some%20people%20online%20are%20experimenting,of%20confidentiality%20and%20safety%20concerns

Lemoine, B. (2022, June 11). *Is LaMDA Sentient? – An Interview.* Retrieved from Medium [Last accessed 23 January 2024]: https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Metz T (2007a) Toward an African moral theory. *The Journal of Political Philosophy 15* (3), 321–341.

Metz T (2007b) Ubuntu as a moral theory: Reply to four critics. *South African Journal of Philosophy 26* (4), 369–387

Metz T (2012) An African theory of moral status: A relational alternative to individualism and holism. *Ethical Theory and Moral Practice 15* (3), 387–402.

Metz T (2021) *A Relational Moral Theory: African Ethics in and beyond the Continent.* Oxford: Oxford University Press.

Mhlambi S (2020) From rationality to Relationality: Ubuntu as an Ethical & Human Rights Framework for artificial intelligence governance. *Carr Centre for Human Rights Policy Harvard Kennedy School, Spring 2020* (9), 1–27.

Millar B and Lopez-Cantero P (2022) Grief, continuing bonds, and unreciprocated love. *The Southern Journal of Philosophy 60*, 413–436.

Newton C (2016, October). *Speak, Memory. When Her Best Friend Died, She Used Artificial Intelligence to Keep Talking to Him.* Retrieved from The Verge [Last accessed 12 March 2024]: https://www.theverge.com/a/luka-artificial-intelligence-memorial-roman-mazurenko-bot

Noddings N (2013). *Caring: A Relational Approach to Ethics and Moral Education*, 2nd Edn. Berkeley and Los Angeles: University of California Press.

Norlock KJ (2017) Real (and) imaginal relationships with the dead. *Journal of Value Inquiry 51* (2), 341–356.

Ogude J (2019) In Ogude J (ed.), *Ubuntu and the Reconstitution of Community.* Indiana University Press, Bloomington, Indiana.

Öhman CJ and Watson D (2019) Are the dead taking over Facebook? A big data approach to the future of death online. *Big Data & Society 6* (1), 1–13.

Ribbens McCarthy J, **Woodthorpe K and Almack K** (2023) The aftermath of death in the continuing lives of the living: Extending 'bereavement' paradigms through family and relational perspectives. *Sociology, 57* (6), 1–25.

Rothpaut JW and Becker K (2007) A literature review of Western bereavement theory: From decathecting to continuing bonds. *The Family Journal 15*(1), 6–15.

Stokes P (2012) Ghosts in the machine: Do the dead live on in Facebook? *Philosophy & Technology 25* (3), 363–379.

Vallor S (2015) Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy & Technology 28* (1), 107–124.

Wells K (2023, June 9) *An Eating Disorders Chatbot Offered Dieting Advice, Raising Fears About AI in Health.* Retrieved from npr [Last accessed March 16 2024]: https://www.npr.org/sections/health-shots/2023/06/08/1180838096/an-eating-disorders-chatbot-offered-dieting-advice-raising-fears-about-ai-in-hea

Wiredu K (1992) Death and the afterlife in African culture. In Wiredu K and Gyekye K (eds.), *Person and Community: Ghanaian Philosophical Studies, I.* Washington, DC: Council for Resaerch in Values and Philosophy, pp. 137–152.

Wiredu K (1992) The moral foundations of an African culture. In Wiredu K and Gyekye K (eds.), *Person and Community: Ghanaian Philosophical Studies, I.* Washington, DC: Council for Research in Values in Philosophy, pp. 193–206.

Worden WJ (1982) *Grief Counseling and Grief Therapy: A Handbook for the Mental Health Practitioner.* New York: Springer.

Worden WJ (1996) *Children and Grief: When a Parent Dies.* New York: Guildford Press.