


ARTICLE

Synthesizing Methuselah: The Question of Artificial Agelessness

Richard B. Gibson 

Institute for Bioethics & Health Humanities, University of Texas Medical Branch, Galveston, TX, USA
Email: rbgibson@utmb.edu

Abstract

As biological organisms, we age and, eventually, die. However, age's deteriorating effects may not be universal. Some theoretical entities, due to their synthetic composition, could exist independently from aging—artificial general intelligence (AGI). With adequate resource access, an AGI could theoretically be ageless and would be, in some sense, immortal. Yet, this need not be inevitable. Designers could imbue AGIs with artificial mortality via an internal shut-off point. The question, though, is, should they? Should researchers curtail an AGI's potentially endless lifespan by deliberately making it mortal? It is this question that this article explores. First, it considers what type of AGI is under discussion before outlining how such beings could be ageless. Then, after clarifying the type of immortality under discussion and arguing that imbuing an AGI with synthetic aging would be person-affecting, the article explores four core conundrums: (i) deliberately causing a morally significant being's death; (ii) immortality's associated harms; (iii) concerns about immortality's unequal assignment; and (iv) the danger of immortal AGI overlords. The article concludes that while prudence requires we create an aging AGI, in the face of the material harm such an action would constitute, this is an insufficient reason to justify doing so.

Keywords: aging; artificial intelligence; death; immortality; moral status

Introduction

The implications of artificial general intelligence (AGI)¹—non-biological consciousnesses with capacities, behaviors, and underlying principles meaningfully comparable to our own—have been debated within and outside academia for decades.² Central to such talks are the normative ramifications such a novel form of being would have and what claims, duties, and rights it might possess or to which it could be subject. Moreover, such discussions typically draw upon anthropological baselines or qualities to inform claims about a non-biological life form's moral status, like sentience,³ consciousness,⁴ cognition,⁵ or behaviors.⁶ This approach, at least intuitively, makes sense. After all, we (hopefully) matter morally. So, according to whatever metrics one employs, if an AGI is significantly and meaningfully comparable, it must also matter.⁷ However, the existing literature overlooks one crucial distinction between biological organisms and AGI: aging and mortality.

With few exceptions,⁸ biological organisms have an in-built expiration date; we die of old age. Even if one avoids life's countless dangers, from murderers to choking on a hotdog and everything in between, while remaining free from fatal pathologies, age inevitably makes corpses of us all. Nevertheless, despite this unwavering conclusion to our lives, we continue living, and, what might be considered even more remarkable, many reproduce, knowing full well that those children, too, will eventually die. While we do what we can to ensure that our progeny live long lives, our reproductive efforts do not invalidate the next generation's perishability. Given that no parent wants to contemplate their child's death, it seems

reasonable that, at least for some, the possibility of their children living an indefinite lifespan would be a tempting proposition. After all, many would consider the option for themselves.⁹

Now, this has never been a possibility. Humanity has never been able to create an ageless lifeform of equal moral worth to ourselves. With the advent of AGI, however, it could be possible to bring into the world a being that, provided it avoided damage and could secure the resources needed for its continued survival, might be ageless. In other words, a form of synthetic life, with cognitive capacities and behaviors comparable to (or exceeding) our own, that, due to its physical and mental constructs being hardware and software rather than biological, is largely free from time's irreversible degrading effects.

This possibility presents an intriguing question. If one can create an ageless AGI, should they? Or is there an argument to be made that aging is a good which we should not deny novel forms of life like AGI? Or, put otherwise, faced with the choice of creating a mortal or immortal AGI, which should we choose?¹⁰ That is the question that this article explores.

The article begins by outlining what type of AGI is under discussion. It charts why such beings could be functionally ageless, what obligations this would impose upon those around them, and why designers and researchers might be tempted to limit an AGI's lifespan. Then, after clarifying what type of immortality is under discussion, the article tackles the issue of person-affecting versus identity-affecting harms, arguing that imbuing an AGI with synthetic aging would constitute the former rather than the latter. It then explores four problems sitting at the core of this conundrum: (i) deliberately causing the death of a morally important being; (ii) the possible harms associated with immortality; (iii) justice concerns of making both immortal and mortal AGIs; and (iv) apprehensions about the creation of immortal AGI overlords. The article concludes that while prudence may demand we create an aging AGI, this is insufficient to justify conferring such harm upon an artificial being.

What Type of AGI Are We Talking About?

Before delving into the topic, some clarification is required as there are various routes to reach the destination of a morally significant, non-biological mind. The classically envisioned form of AGI—embodied in popular culture by *Star Trek's* Data, *Westworld's* Hosts, *Tron's* virtual intelligence, and *The Matrix's* machines and agents—is not the only way such consciousnesses might exist. A quick outline of two forms of AGI and AGI-adjacent entities is needed as not all require equal consideration.

Instead of creating an AGI from scratch, an artificial consciousness could emerge via the transference or replication of an existing biological mind into an artificial construct, otherwise known as mind uploading. As Damien Broderick outlines,

*[W]e could become machines while remaining ourselves, physically transferring the structure of our minds into capacious computer programs that generate thought and the quality of minds when they are run. Uploads would live in vivid virtual realities fitted to the needs of their simulated minds, while remaining in touch with the external world.*¹¹

The processes by which this might occur are unclear. For example, we do not know whether uploading would leave an unconscious biological shell in its wake, destroy the brain, or copy a mind, twinning the organic consciousness with an artificial counterpart. While the exact mechanisms remain vague, we can infer that such an artificial consciousness would not emerge into the world anew. The point of such uploading is to give the digital entity a matching identity to that of the modeled biological mind.¹² If this were not the case, then it would not satisfy the goal of mind uploading, that is, creating a digital version of an existing entity. So, this article does not refer to mind uploads when discussing AGI. Instead, it is chiefly concerned with creating a new, distinct consciousness.

Another way an artificial consciousness may emerge is not by existing independently of biology but in tandem with it as a cyborg. This consciousness would not necessarily be entirely artificial as it would incorporate a degree of biological consciousness while non-biological components generate aspects of that mind. What percentage of such a composite psyche would arise from artificial elements and what

from biological ones would vary. Researchers might augment existing organic components with neurological implants (like *Star Trek's* Borg), or the inverse might be true, where scientists alter an artificial entity by incorporating biological material (like *Battlestar Galactica's* Hybrids). As with minds emerging from uploading, the entity resulting from such a union is not considered here. Such cyborgs may not be entirely new consciousnesses as they would incorporate biological materials into their artificial systems, and even if such a being were a blank slate, the biological materials would confer aging, precluding it from the agelessness discussion.

Instead, the entity under question emerges into the world as a distinctive, subjectively new consciousness lacking memories or prior states of existence. It is analogous to a newborn baby entering the world as a blank slate lacking preexisting frameworks or subjective perspectives; it is untangled from any preceding identity and is an artificial individual whose qualities—whatever they may be—entitle it to human-equivalent moral worth. Unlike a human, however, the consciousness here emerges not from neuronal firing but from binary or quantum computer code processed by hardware similar to, or at least recognizable as developed from, existing hardware. It is *Star Trek's* Data upon his first activation rather than *Robocop's* Alex Murphy.

The above description is not a definitive account of what type of AGI may or will exist. Indeed, at this relatively early developmental stage, any claims of certainty regarding AGI embodiment should be viewed with skepticism. Instead, the account provided functions to give form to the entity under consideration. It is an example, not a prediction. Additionally, while the provided account is tethered to hardware—much like consciousness is to the nervous system—this need not be necessary. Software and hardware's propensity for interconnectivity means that an AGI may not have any fixed physical presence, instead existing in the web of connections that make up the Internet and similar networks.

The critical point for this article, one existing in separation from the specific form of AGI that is designed and brought into the world, concerns a trait which, regardless of an AGI's exact condition, is likely to be shared by all of them: its capacity for longevity.

Well-designed software of the type generating AGI may run indefinitely. It seems possible, providing that it is free from catastrophic design flaws, that an AGI could operate without the degenerative issues that affect biological entities (forgetfulness, confusion, reduced comprehension, physical weakness), or at least such temporally induced degradation is not an intrinsic trait of such an entity. In other words, if designed correctly, an artificial mind could keep working. Admittedly, the hardware upon which such a mind runs would need maintenance for this to occur. If that hardware becomes damaged from an accident, general wear and tear, or deliberate sabotage, the AGI will likely experience difficulties. Unlike a biological brain, however, such hardware could be replaced. Of course, this would be far more of an issue in cases where an AGI remains tethered to specific hardware. Cloud-based AGIs would only be affected by catastrophic, systematic hardware failures; anything less, and the AGI could transfer to undamaged equipment.

All this is to say that it seems plausible (or even potentially likely) that an AGI, given adequate resource access and freedom from damage, could function and, in parlance appropriate here, live free from time's deleterious effects. In other words, when creating an AGI, we could create a consciousness requiring equal moral consideration of degree and type as a person. However, unlike anyone who does or has existed, such an entity would be free from aging's fatal consequences.

This potential, however, complicates the already complicated ethical map of whether AGI creation is morally acceptable. As I have argued previously,¹³ creating an AGI also creates duties for those nearby. Most notable would be the obligation to supply the AGI with essential resources, like electricity, software updates, and hardware repairs. Similarly to how one is obliged to prevent someone from dying of thirst or starvation where possible, the same would be true for an AGI possessing moral worth. If an entity is morally significant, then its synthetic rather than biological constitution should be irrelevant—what Nick Bostrom and Eliezer Yudkowsky call the Principle of Substrate Non-Discrimination.¹⁴ Thus, a duty to save one translates to the other—we should prevent an AGI from “starving” to death from a lack of energy or similar other resources, just as with a person.

Nevertheless, while such obligations fade regarding humans, given that they ultimately die, the same cannot be said for an AGI. Provided it continues living, there is a continuing obligation to furnish it with life-sustaining resources. As I have argued in past work,

[I]f there is an obligation to ensure that a machine consciousness has the materials required to continue its existence, such as power or software updates because accidentally or deliberately failing to do so intrinsically wrongs the artificial consciousness, then this obligation could be equally ageless. For as long as the artificial consciousness “lives,” such an obligation exists. And if the machine consciousness fails to expire from old age and is functionally immortal, so too would be such an obligation.¹⁵

In response to such a possibility, one might consider ways to avoid binding ourselves in such a manner. That is, deliberately designing AGIs sans the possibility of placing an eternal obligation on us to provide them with life-sustaining resources. The appeal of this is evident. We would gain AGI’s associative benefits without acquiring additional, likely demanding, obligations.

One way to do this would be to design AGIs more like us regarding our relationship with death—even with access to life-sustaining resources, they would still perish; in other words, to create an aging, mortal AGI. While the specifics of how designers could achieve this are outside this article’s scope, broadly speaking, it would be a case of embedding within an AGI’s core code instructions to, at some point, irreversibly shut down. Such a point could derive from the AGI’s activation date, the number of functions it has completed since coming online, or even arbitrarily. What matters is that regardless of the AGI’s efforts, the point of cessation is inevitable; that, just like us, regardless of how many actions we take to stave off death, eventually, we die.

However, should those creating such an AGI knowingly and deliberately limit its lifespan? If designers can create a functionally ageless AGI but instead create one with a limited lifetime, do they commit some moral transgression?

A Note on Agelessness and Immortality

Having outlined what sort of AGI is under discussion here, and before delving into the question at hand, a quick acknowledgment regarding the relationship between agelessness and immortality is required.

One might consider agelessness synonymous with immortality. This, however, is not the case, and classical and contemporary literature abounds with examples where entities possess one or another. That is, they can age but not die or die but not age. In Greek mythology, for instance, Zeus awarded Tithonus, a prince of Troy and lover of Eros, immortality. However, while Zeus granted this wish, he (accidentally or deliberately) failed to make Tithonus ageless. Thus, as time passed, Tithonus grew increasingly wizened as he aged yet could not die, eventually lamenting his lost mortality.¹⁶ Inversely, Adventure Time’s Princess Bubblegum, after initially appearing young and growing into early adulthood, ceases to age in any conventional sense and remains physically 18 while over 800 years old. However, she is not truly immortal, as she can die from an accident or attack.¹⁷ These examples show that conflating agelessness with immortality can give rise to conceptual imprecision. We can use the term in one way while unintentionally including or excluding another. Instead, immortality comes in degrees. We might consider some entities immortal because they possess an extended but not indefinite lifespan (possibly enabled via scientific or mystical interventions), others because they are ageless but can still die, and some may possess true immortality because they cannot perish under any circumstance.¹⁸

What is under discussion here is the second of these categories. Not that an AGI might have a longer lifespan compared to other similar entities; indeed, it would be hard to make sense of this form of immortality in this context as it requires a non-existent comparable baseline against which to understand the property of extendedness. Nor that they cannot die regardless of what befalls them, as this is an impossibility for any entity adhering to the principle of entropy as understood in the second law of thermodynamics.¹⁹

Instead, what is of concern here is immortality in the sense of an entity existing independent of aging's deleterious and eventually fatal effects. It is this form of immortality that an AGI has the potential to obtain.²⁰ Or, to put it more accurately, a quality that an AGI may or may not possess according to the intentions of those designing it.

So, if a researcher can choose between creating an ageless or aging AI, do they commit a wrong if they pick the latter?

Selection and the Non-Identity Problem

When considering decisions over whom to create and the conditions upon which those that exist can harm the yet-to-exist, one invariably evokes Derek Parfit's non-identity problem.²¹ According to the problem, it is nearly impossible to harm an individual through the actions one makes before that individual exists, even when said actions negatively impact that eventual person, provided their life is worth living. This is because such decisions do not harm a singular individual but alter the individual that comes into existence. Thus, one cannot be harmed by the outcomes of decisions that, if not taken, would result in that person's non-existence.

This problem often arises in embryo selection debates where one embryo is implanted instead of another. For example, in selecting between a hearing and deaf embryo, one does not harm the eventual person by selecting the latter, as that individual would not exist if one had selected the former.²² However, if one alters the genetics of a singular embryo before implantation, this does not fall foul of the non-identity problem as the individual's traits are altered, but they themselves remain. Thus, one could argue that the eventual person has been harmed in situations where one alters an embryo's genetics to shift from hearing's presence to its absence. While bioethicists have explored this in biological contexts, it also seemingly applies in AGI cases.

At its core, the issues remain. Instead of considering the alteration of genetic material, however, we would be examining the ethics of altering or adding to an AGI's coding. In this context, the question regarding the non-identity problem is: Would changes made to such an AGI before its activation be identity or person-affecting?

While the comparison between selecting an AGI and an embryo provides some insight, it also encounters issues as the two entities are dissimilar concerning their relationship with the physical world. Nature tethers the identity under question in the classical version of the non-identity problem to a physical structure—that is, the body. So, when discussing embryo implantation, the focus is on the embryo's physicality; a different physicality constitutes a distinct identity. In instances of AGI, however, a comparable physical body is absent. While such an AGI would still need hardware to function, this hardware would not be synonymous with the AGI. The AGI could be transferred to different hardware while remaining the same entity as its coding, which gives rise to consciousness, would remain intact. This is not true of biological organisms, where one's intelligence and body are inseparable.²³

This issue's intricacies run far, and this article lacks the full scope to delve into them. It seems reasonable, however, that some degree of alteration or addition to an AGI's code could be made without it necessitating an identity-affecting change—one can change an AGI's traits without changing the AGI.

The coding resulting in an AGI's emergence will undoubtedly be complex. Nevertheless, there would likely be room for code alterations that, while impacting an AGI, do not generate a wholesale new entity. For illustration, designers could change the color of an AGI's display interface via a coding alteration without altering the nature of that AGI; it would be a trait alteration.²⁴ Alternatively, changes to an AGI's functions could be made not by altering existing code but by supplementing it, providing additional functionalities to an already completed, unactivated AGI. This would be akin to downloadable content for existing programs.

All this is to say that alterations to an AGI's code need not be identity-affecting and could, dependent on how substantive any changes are and whether they are substitutions or additions, equate to person-effective changes. The AGI would not become anew by such alterations but receive new characteristics. These could be as innocuous as changing the interface color or as significant as conferring mortality.²⁵

While such an assumption seems reasonable, it is certainly not guaranteed; the conferring of immortality might be so radical as to constitute a person-affecting change. As a result, it would mean that mortality would not change the traits of a single AGI but the AGI itself, resulting in a different entity being developed. This, in turn, would mean that discussions about whether researchers should confer mortality upon a single AGI are nonsensical as there is no single AGI at the centre of such a discussion – there is a mortal AGI and an immortal AGI. Thus, we are left with the question of which to bring into the world.

Yet, at least for this article, such a concession need not be a disaster. Instead, it would shift the discussion from one about conferring mortality onto a single entity to one about choosing which entity, out of a choice of two, researchers should create: a mortal *or* immortal AGI. This, in turn, would evoke questions regarding procreative beneficence, which, defined by Julian Savulescu, is the principle that

*[c]ouples (or single reproducers) should select the child, of the possible children they could have, who is expected to have the best life, or at least as good a life as the others, based on the relevant, available information.*²⁶

While Savulescu frames this definition in terms of parent(s) selecting for their children, just as with the non-identity problem, in theory, the principle applies to questions concerning which non-biological entities researchers should create. Regarding the work contained here, the principle of procreative beneficence would demand that designers create those they expect to have the best life. In other words, they would choose between making a mortal or immortal AGI, each entity being numerically distinct based on which would have a better life.

The topics addressed in this article could inform such a discussion because whether postponing death is good, whether immortality is inherently harmful, and whether discrepancies in mortality's assignment would all factor into evaluating whether an AGI would live the best life possible. The only topic that might not be as relevant to this AGI (im)mortality approach is the risks of an immortal overlord.

Nevertheless, as already noted, this article assumes that the bestowment of mortality upon an AGI would result in person-affecting changes.²⁷

The Ethics of AGI (Im)mortality

Working from the assumption that imbuing an AGI, which possesses intrinsic moral value, with mortality is person-affecting, we can explore whether researchers commit a wrong by creating a mortal rather than immortal synthetic consciousness. We start with a somewhat uncontentious claim.

Postponing Death Is Good

We typically avoid death, and evolution and society have baked into our very foundation the desire to frustrate the destruction of ourselves and others. As Geoffrey Miller writes,

*There is, of course, no way to escape the hardwired fears and reactions that motivate humans to avoid death. Suffocate me, and I'll struggle. Shoot me, and I'll scream. The brain stem and amygdala will always do their job of struggling to preserve one's life at any cost.*²⁸

While exceptions exist, such as when one wishes to die to avoid suffering or save the life of another, as a general rule, it seems reasonable to assert that we prefer life over death.

Furthermore, we usually portray actions, innovations, and ideas promoting life positively. We praise those who work toward people's betterment and vilify those who set out to harm deliberately and, at the extreme, kill others. This attitude is one reason why discussions around physician-assisted suicide, for example, are problematic—they involve ending a life, which rubs against the grain of collective moral wisdom. Even in philosophical thought experiments, the intuition that we should avoid death is

commonplace and central. Think Peter Singer's drowning child,²⁹ Philippa Foot's trolley problem,³⁰ or Judith Jarvis Thomson's famous violinist.³¹ While each explores a different dilemma, they hold, as a foundational premise, death's undesirability. Indeed, according to Albert Camus, philosophy's very purpose is to reconcile our relationship with death; to understand why we should prefer life and almost always resist its alternative.³² All this is to say, all things being equal, life is good, and death is, comparatively, bad.³³

If one believes that the good should be promoted and the bad impeded, it follows that we should take action to promote life and thwart death. However, the latter is impossible for us mere mortals. While clinicians' efforts may send cancers into remission and stem internal bleeding with surgery and sutures, death can never be defeated, only postponed. Regardless, life still holds value; saving lives remains praiseworthy, despite such acts' ultimate futility, and killing others remains abhorrent even though one's victim invariably dies eventually. While commonly accepted in cases involving biological organisms, this claim would also seemingly apply to entities comprising other material substrates with comparable moral worth, like AGIs.

Working from the baseline that preventing death is good, it seems that artificial aging should not be afforded to AGIs because doing so would contravene this general death prevention principle. Synthetic mortality would limit the amount of life an AGI has by reducing the number of days it lives and accelerating the point at which death occurs. To draw a somewhat imprecise analogy, it is akin to altering an embryo's genetics so that, rather than living for 80 years, it lives for only 40—a *prima facie* wrong.

The harm in AGI cases, however, is more significant than in that of a biological entity as death was always a factor in the human's future; it was a question of when, not if. This inevitability is not the case for the AGI. Ageless immortality was an option for it, which a designer actively prevents the AGI from possessing when they give it lifespan-limiting software. Thus, the harm is significantly more severe in this case because it is no longer a matter of reducing the number of years a being has on this earth from one limit to another but bestowing a previously non-existent limitation. It is not a case of further limiting the limited but of limiting the limitless. As harmful as it is to cause a person's early death, it is worse to do so for a comparable entity for whom any death is premature.

Ultimately, if we believe that prematurely ending a life is morally wrong, then ending the life of an immortal AGI by making it age and eventually die is, at the very least, equally wrong. Thus, it is an action from which AGI designers should refrain.

The Possible Harms of Immortality

Despite the seeming objective goodness of more life, it does not necessarily follow that eternal life is unendingly good. One can have too much of a good thing, and immortality could be an endless drudgery. So, by imbuing an AGI with aging and, thus, mortality, one might save them from a fate worse than the alternative. Indeed, such a position has been explored and defended in the context of human immortality from several avenues.

First, there is the argument that death's inevitability gives life meaning and that the former's elimination does not provide a boundless pasture to roam but destroys the thing that makes life's meadow lush. Immortality would increase life's quantity while negating its quality, rendering it meaningless. As Viktor Frankl writes,

*For what would our lives be like if they were not finite in time, but infinite? If we were immortal, we could legitimately postpone every action forever. It would be of no consequence whether or not we did a thing now; every act might just as well be done tomorrow or the day after or a year from now or ten years hence. But in the face of death as absolute finis to our future and boundary to our possibilities, we are under the imperative of utilizing our lifetimes to the utmost, not letting the singular opportunities... pass by unused.*³⁴

According to Frankl, knowing one has endless days to complete tasks means infinite time to procrastinate. Immortality would eliminate a key driver in our lives—that we have a set limit to do what we want

—and without this driver, we would lose the motivation to do anything. The fact that, as mortal beings, our window of opportunity is forever closing provides us with the motivation to achieve goals and experiences before it is too late. It would also mean that once-in-a-lifetime opportunities would lose their rarity, devaluing some uniquely valuable experiences. In the AGI context, granting them aging would mean providing them with an otherwise absent motivation. It would provide the drive to complete the tasks and goals they might otherwise defer.

It is unclear, however, whether AGIs would require such motivation as they may not be procrastinators. Contemporary software does not exhibit procrastinatory tendencies, and while an AGI will significantly differ from existing programs, it would be a descendant. It is unclear why it would need to procrastinate or when such a trait would emerge. If we do not program it with such a dilatory drive (and why would we?), this desire to put tasks off would seemingly be absent. Thus, there would be no need for death as a motivator because the AGI would already operate free from the propensity to postpone. It would not procrastinate; it would just do.

Additionally, even if one assumes that an AGI would have or need to procrastinate, this does not immediately lead to the need for death as a motivator. As John Martin Fischer and Benjamin Mitchell-Yellin,³⁵ and John Harris note,³⁶ even for an immortal, some things would still be time-dependent, such as securing resources or interacting with mortals. Just because an AGI is eternal does not mean that its surroundings are, and to operate in an ever-changing world means that external change still makes some things time-dependent. So, conferring upon an AGI synthetic aging may not fill a need but instead limit it in a way that provides, at least in this regard, no absentee motivator.

Second, there looms the specter of boredom. As Bernard Williams argues, immortals would have enough time to overcome all obstacles, complete all tasks, and exhaust the catalog of experiences. Absent the possibility of new challenges, life would become dull. Existence would be an unending clip show where unfolding experiences are mere echoes of the past. As Williams illustrates, drawing upon the example of Elina Makropulos, the subject of a twentieth-century play who maintains her immortality by a potion,

Her trouble was, it seems, boredom: a boredom connected with the fact that everything that could happen and make sense to one particular human being of 42 had already happened to her. Or, rather, all the sorts of things that could make sense to one woman of a certain character...³⁷

For Williams, Elina's longevity creates the lack of what he terms categorical desires—reasons for one to continue living.³⁸ The boredom permeating Elina's life is all-encompassing, and regardless of her attempts to revitalize her existence, she comes to wish for oblivion. To put it indelicately, she is bored to death. Similar arguments have been advanced by Shelly Kagan,³⁹ Martha Nussbaum,⁴⁰ Stephen Cave,⁴¹ Samuel Scheffler,⁴² and Todd May.⁴³

Like the never-ending procrastination argument, however, whether an AGI would suffer from the absence of a categorical desire like Williams envisions Elina Makropulos as undergoing is unclear. It could be that an AGI cannot become bored because we do not design it with the ability. Alternatively, even if it can, it is unlikely to because it has practically infinite realms of virtual reality and the material world with which to interact. When bored with the material world, it can enter the virtual one and create new realities to explore and alter. Thus, the concern about an AGI becoming bored with existence seems, at least at a cursory glance, to be unlikely.

For the sake of argument, though, let us assume that an AGI can indeed become bored. Even then, there is reason to believe immortality might not become intolerably dull. Instead, as Thomas Nagel argues, it might be fun:

Couldn't they [immortal lives] be composed of an endless sequence of quests, undertakings, and discoveries, including successes and failures? Humans are amazingly adaptable, and have developed many forms of life and value in their history so far, in response to changing material circumstances. I am not persuaded that the essential role of mortality in shaping the meaning we find in our actual

*lives implies that earthly immortality would not be a good thing. If medical science ever finds a way to turn off the ageing process, I suspect we would manage.*⁴⁴

While focusing on humans immortalized by biomedical technologies, Nagel's point would apply to nonhumans. Engaging with the virtual and material worlds, where one could learn, play, and grow, might remain engrossing for decades and even millennia. Just because some people tend to get bored by repeating their daily lives does not mean that all do. Also, as already noted, immortality does not bring a static existence; while one may be eternal, one's experiences would shift alongside the changes in the world.

The final harm of immortality that this article will discuss concerns forgetfulness. Human memory is flawed. Not only do we forget things, but also those we remember are often recollected with inaccuracies or are entirely fabricated.⁴⁵ Over a lifetime, we fail to recall countless things because we were not paying attention, they happened long ago, or we develop pathologies that inhibit our recollection. So, even if Williams is right that immortality might exhaust our categorical desires, unless we remember satiating such desires, we can re-satisfy them. As Christopher Belshaw suggests,

*My boredom at seeing Hamlet for the twentieth time depends not just on repetition, and my having seen it those nineteen times before, but also, to some considerable extent, on remembering, and in some detail, what I have seen. Wipe out the memories and I see it each time as something new... Obliterate memories of the past and there is no reason not to go on, and in the same vein.*⁴⁶

In other words, our forgetfulness might remedy Williams's boredom problem, as one cannot be bored by repetition if one is unaware that repetition exists. However, as David Blumenfeld and Roy W. Perrett note,⁴⁷ the desirability of such repetition over an immortal lifespan is debatable. Nevertheless, desirable or not, for one to become bored with one's life because of the lack of new opportunities presupposes that one possesses an account of previously seized opportunities. Without this recollective fidelity, Williams's curse of immortal boredom seems inconsequential.

Recollective failure, while a very human characteristic, is not something an AGI may suffer as software has an uncanny ability to recall accessible data. Computers can retrieve files initially saved years or decades ago with relative ease and high fidelity. An AGI that runs on more advanced versions of existing software and hardware will likely share this recollection capacity. It will have an eidetic memory, capable of remembering any event that happened to it, meaning that, regarding Williams's categorical desire concern, an AGI is unlikely, perhaps even unable, to forget those desires it satisfied years, decades, or even millennia prior. While a human will forget and be able to complete them seemingly anew, for the AGI, once a desire is satisfied, it seemingly cannot be re-satisfied.

Unless, that is, the AGI can delete memories like we can delete files from a computer's drive. If possible, an AGI could theoretically select a desire it believes it might want to re-satisfy, erase the files associated with the desire's existing satisfaction, and then retackle the task, challenge, or goal anew. This venture could be tricky to do in practice, however, as the selective deletion of memories can be challenging to achieve without compromising broader system stability (think of the damage caused when accidentally deleting a system-critical driver from your laptop). For an AGI, whose architecture would be far more complex than anything existing today, deleting memories might risk compromising the AGI itself. However, this risk is far from guaranteed.

Whether immortality could be a harm we should save an AGI from is tricky to ascertain. Given the differences in AGI and human motivations and memory, direct comparisons between the two groups provide only limited insights. Nevertheless, it seems plausible that an AGI need not possess the all-too-human (or at least biological) traits of procrastination and forgetfulness. Indeed, while an AGI could be designed with these qualities, thereby making it more human-like, it is unclear why any researcher would do this as it would impede an AGI's functionality.

Additionally, even if immortality proves harmful, which there is no guarantee it is, this does not mean that an AGI *must* continue living. As the AGI is ageless rather than invulnerable, it could, if desired, end its existence by self-destructing or, to put it in human terms, commit suicide. Whether it would be able to

do this by itself is unclear, depending as it is on the AGI's constitution and capabilities. However, in theory, it could ask others to assist it in ending its life. Indeed, the question of whether AGIs could utilize existing legal methods via which humans can gain assistance in ending their lives has already been posed, most notably by Isra Black.⁴⁸

Ultimately, while there might be reasons for humans to shun immortality, these fail to provide reasons to deny AGIs such longevity.

What Is Good for the Electric Goose Is Good for the Electric Gander

The preceding discussion explored whether researchers should imbue an AGI with artificial aging based on the desire to prevent death and immortality's potential harm. However, those responsible for creating AGI are unlikely to stop at one. Once one AGI exists, more are likely to follow as the societal and economic advantages of such a new life form will increase alongside their population. This likelihood brings additional issues that a single AGI's existence does not, such as whether (im)mortality must be assigned universally or selectively.

There are seemingly three answers to this question. Either (i) no AGIs are afforded aging; (ii) all AGIs are afforded aging; or (iii) some AGIs are afforded aging while others are not. Each option, though, brings complications.

If it is impermissible to make a single AGI with the capacity to age and die because doing so would limit that individual's life and, ultimately, kill them when they need not have died, then this would also apply across all AGIs. As such, researchers would need to refrain from conferring aging to all AGIs. This stance, in turn, would allow the steady growth of a nonorganic population. This population, though, might be vulnerable to one of the issues central to the human-immortality discussion—overpopulation.⁴⁹

There are already high demands on many resources that AGIs would need to survive, such as cobalt, silicon, and electrical energy. Increased numbers of AGI would need increased access to resources, and failure to get this access could result in starvation equivalent harm. Even if resources were available in the amounts needed, assignment and distribution would be unlike any challenge humanity's collective societies have faced. The many systems put in place over the centuries to enable societies' growth have been created and enacted based on the knowledge that those using them die and future generations replace them. Creating a system with nonhuman and immortal users would radically depart from this precedent and likely present countless unforeseeable challenges.⁵⁰ Thus, death might be necessary to continue the circle of life, which, while mandatory for biological organisms and facilitating new generations' access to resources, might be optional but no less vital for nonorganic life.

Inversely, if one agrees that immortality, for whatever reason, is undesirable and that researchers would be obligated to integrate aging into the systems of all AGIs, this, too, would bring problems. These include deciding at what point each AGI should "die." Would each AGI have the same lifespan, or would their perishability vary according to some deliberate or arbitrary factor? When facing death, should these AGIs be able to replicate themselves anew, creating a form of reincarnation or perhaps progeny? One could also ask questions about the motivating factor for AGI mortality in that it might prove advantageous for humans to have an AGI that dies (such as when they become obsolete). However, denying them immortality for *our* benefit appears inherently wrong. It would reduce their existence, indeed their mortality, down to a means to our ends, treating them as objects rather than moral entities.

One further point regarding the wholesale mortalization (for lack of a better word) of AGIs that requires acknowledgment is the possibility that it would be akin to, perhaps identical to, genocide. Article 2 of the 1948 Genocide Convention defines genocide as follows:

*[A]ny of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such: (a) Killing members of the group; (b) Causing serious bodily or mental harm to members of the group; (c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part; (d) Imposing measures intended to prevent births within the group; (e) Forcibly transferring children of the group to another group.*⁵¹

It is not inconceivable to see researchers imbuing aging onto what would otherwise have been ageless AGIs as, in some manner of speaking, killing them—all be it via a delayed method. After all, if an entity was not going to die, but then an individual knowingly and deliberately takes actions that cause that entity's death, it seems reasonable to assert that they have killed that entity. This would be impermissible in singular cases, but when it comes to the deliberate ending of lives across all AGIs, we face the prospect of killing not a single nonorganic life form but an entire group. Now, this article is conscious of the potential implications of such a claim and does not wish to throw such significant charges around without appreciating their meaning. Nevertheless, while it is unclear whether AGIs could qualify as a group under the Convention, the potential is there, and given such stakes, researchers, academics, designers, and lawyers must consider this possible infraction.

What both of these blanket approaches to immortality have in common, which is a benefit, is their uniformity; they treat like for like. Moreover, if immortality should be denied or afforded to one AGI, for whatever reason one finds convincing, then, seemingly, this should apply to all AGIs. Nevertheless, as indicated, they are not the only options, as researchers might conclude that they should make some AGIs mortal while others should remain immortal. This might be for instrumental reasons (some AGIs might fulfill their roles better if they perish regularly or not at all), market factors (developers could sell perishable AGIs for less than immortal ones), or countless other reasons. Such an uneven mortality distribution, however, is problematic.

First and foremost is the fact that researchers would, to use emotive language, be deliberately responsible for the deaths of some AGIs but not others. Consigning some AGIs to a limited lifespan but not others would unevenly apply the benefits and harms of immortality and mortality. If AGIs are comparable, this would be a significant, perhaps even unparalleled, injustice. Consider if someone was in a position of power to do this in human cases. The uneven wielding of such power would constitute a great injustice; all things being equal, either we all deserve immortality or none do. The same would apply not just to humans but also to AGIs, regardless of their non-biological constitution.

Second, even if one could justify the uneven application of aging, there would emerge issues in how one decides the distribution pattern—why would some AGIs qualify for mortality and others not? This issue becomes even more tricky when we remember that the point of aging's introduction comes before the entity in question exists. Thus, it cannot be done according to merit or performance, as each AGI would not have had the chance to demonstrate why it should or should not be made mortal. Economic factors might dictate mortality's assignment. If more people want to pay for an immortal AGI than a mortal one, market factors will lead researchers to create more of the former. However, deciding at what point morally significant beings should or should not die according to the financial whims of those paying for them seems patently immoral (not to mention, if AGIs are comparable to us, perhaps even a form of slavery).⁵²

Finally, leaving some AGIs with their immortality intact while others have it removed would result in a parallel population of similar, but not identical, beings; some living endlessly, others doomed to death. Depending on these novel beings' cognitive composition, this inequality may result in resentment and direct harm. After all, inequality in human populations has historically led to societal turbulence, poorer life outcomes, and, in extreme cases, violent revolution.⁵³ For a demographic of an AGI population to face mortality while another group lives free from such concerns would be a considerable, perhaps unrivaled, inequality. This disparity, in turn, might have severe implications for those individuals responsible for the discriminate assignment of mortality—the humans who created the AGIs.

The Immortal AI Overlord

While death is typically best avoided, there are some cases where its presence is a blessing. Not to put too fine a point on it, but people and the planet are sometimes better off when evil people perish. Without death, some of history's greatest monsters would still be with us today; their crimes could have continued for decades or centuries beyond their original cutoff point without the Reaper's intervention, and the

control that such villains could have amassed and consolidated would have meant that their terrible powers could have grown even more. So, while death may be bad, certain people's death is arguably good.

Immortality, however, would remove such relief from tyranny, torment, and destruction. An ageless dictator could continue to exercise complete control over their country and populace without fear of incurring the wrath of the oppressed. They would even find relief from the more mundane ways that oppressors pass from this life to the next, like accidents or pathology. For those living under such rule, no death could mean no reprieve—an immortal dictator can dictate in perpetuity.

This nightmare would be just as true in cases of AGI as it would for people. As often illustrated in dystopian narratives like *The Terminator*, *I Robot*, *The Matrix*, and countless others, AGI could theoretically exercise its computational prowess to attack human or, more broadly, biological life.⁵⁴ This attack need not be motivated by malicious sentiments. As Bostrom notes with his paperclip-producing AI apocalypse example, an AGI could commit what we might think of as unbelievable horrors due to programming that, at least on the surface, was designed to promote purely benevolent (or innocuous) actions.⁵⁵

Without the prospect of death, an immortal AGI, one which decides, for whatever reason, to take control, be that as a malevolent or benevolent dictator, could dominate those upon whom it exercises power without fear of its reign ending. Or, at least, without fear that its reign will end due to the same biological clock under which existing dictators live.

This possibility presents a compelling argument for imposing mortality on all AGIs. By subjecting them to an inevitable and irreversible aging process, we introduce death into what would otherwise be an immortal existence. This measure ensures that the AGI's existence will ultimately end, regardless of how uncontrollable or harmful it becomes. It is essential to acknowledge that this doomsday scenario appears unlikely. It is reasonable to expect a litany of safeguards to be implemented during the creation of AGI to prevent such an outcome. Furthermore, even if these safeguards fail, there is no guarantee that an AGI would immediately embark on destructive behavior, as it may not be interested in dominating endeavors.

Nevertheless, this cannot be assumed. The risk does exist, and, despite all efforts, safeguards can fail. Thus, redundancies are always welcome when the stakes are high.⁵⁶ Only having mortal AGIs would mean that, in this worst-case scenario, such horrors might have an end date. Of course, this is not guaranteed. The AGI could remove such a limitation from its programming, design new AGIs that are immortal, or replicate itself when its death is imminent. However, such synthetic mortality would at least be one further safeguard upon which we could pin our hopes if the worst comes to worst.⁵⁷

Conclusion

Unlike us mere mortals, AGIs may possess the capacity to be long-lived. While humans succumb to age and death, such a fate need not wait for an artificial entity. The fact that their consciousness emerges and runs on hardware rather than biology means that, unlike us, provided they avoid damage and can secure essential resources, they could exist for decades, centuries, or perhaps even longer. This potential raises questions about how we would relate with a morally comparable entity that, while not only being fundamentally different from us in construction and capabilities, also differs according to the scale of its lifetime. To put it mildly, altering the globe's philosophical, legal, economic, political, and societal systems to make space for the advent of AGI will be challenging. However, to do all this, knowing that such entities' lives will play out according to far grander scales than our own makes the challenge even more significant.

In the face of such a significant problem, researchers could reject the creation of an ageless AGI and, instead, design them with a built-in, unavoidable, and undetectable death clock—a limiter on how long that AGI will live which, regardless of its actions or efforts will, one day, result in its demise. On the surface, this sounds unforgivable, as we should not kill entities possessing human-equivalent moral worth. Nevertheless, just as there are arguments against the quest for human immortality, there too are ones casting doubt on whether an AGI's immortality would be a blessing.

This article has given a brief overview of some of the arguments that might inform researchers' thinking regarding whether they should imbue an AGI with mortality. After clarifying what sort of entity was under discussion and how aging could be person- rather than identity-affecting, it considered four broad arguments: (i) deliberately causing the death of a morally important being; (ii) the possible harms associated with immortality; (iii) justice concerns of making both immortal and mortal AGIs; and (iv) apprehensions about the creation of immortal AI overlords.

Ultimately, while there are certainly reasons for limiting the lifespan of an AGI, be that singular or plural, the strongest of these reasons appear to be more for our benefit than that of the artificial organism. To argue that their lifespans should be reduced in length because it might prove economically beneficial for us or may provide some safeguard against a rogue AGI appears to deny those synthetic beings the very thing that makes them (ironically) unique—their humanity.

The question is not whether AGI immortality would be in our interest but whether it is in theirs. At least from what this article has outlined, the answer would be no.

Competing Interest. The author declares none.

Notes

1. This very terminology and definition is debatable. Alternative terms, such as artificial intelligence, strong artificial intelligence, or artificial superintelligence, are also contenders for usage; each emphasizing different qualities or capacities. As the most commonly employed term, AGI is the one chosen here.
2. Asimov I. I. *Robot*. New York: Gnome Press; 1950; Good IJ. Speculations concerning the first ultraintelligent machine. *Advances in Computers* 1966;6:31–88; Lloyd D. Frankenstein's children: Artificial intelligence and human value. *Metaphilosophy* 1985;16(4):307–18; Yudkowsky E. Artificial intelligence as a positive and negative factor in global risk. In: Bostrom N, Cirkovic M, eds. *Global Catastrophic Risks*. Oxford: Oxford University Press; 2008:308–45; Erman E, Furendal M. The global governance of artificial intelligence: Some normative concerns. *Moral Philosophy and Politics* 2022;9(2):267–91.
3. Bostrom N, Yudowsky E. The ethics of artificial intelligence. In: Frankish K, Ramsey WM, eds. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press; 2014:316–34; DeGrazia D. Robots with moral status? *Perspectives in Biological Medicine* 2022;65(1):73–88; Gibert M, Martin D. In search of the moral status of AI: Why sentience is a strong argument. *AI & Society* 2022;37:319–30; Sparrow R. The Turing triage test. *Ethics and Information Technology* 2004;6:203–13.
4. Graziano MSA. *Rethinking Consciousness: A Scientific Theory of Subjective Experience*. New York: W.W. Norton & Company; 2019; Levy D. The ethical treatment of artificially conscious robots. *International Journal of Social Robots* 2009;1:209–16; Mosakas K. On the moral status of social robots: Considering the consciousness criterion. *AI & Society* 2021 36:429–43; Chella A, Pipitone A, Morin A, Racy F. Developing self-awareness in robots via inner speech. *Frontiers in Robotics and AI* 2020;7:16.
5. Neely EL. Machine and the moral community. *Philosophy & Technology* 2013;27:97–111; Sinnott-Armstrong W, Conitzer V. How much moral status could artificial intelligence ever achieve?. In: Clarke S, Zohny H, Savulescu J, eds. *Rethinking Moral Status*. Oxford: Oxford University Press; 2021:269–89.
6. Danaher J. Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics* 2020;26:2023–49.
7. Some refute this assumption, arguing that artificial intelligence cannot hold moral status or be considered a moral agent. See: Hakli R, Mäkelä P. Moral responsibility of robots and hybrid agents. *The Monist* 2019;102(2):259275; Constantinescu M, Crisp R. Can robotic AI systems be virtuous and

why does this matter? *International Journal of Social Robotics* 2022;14(6):1547–57. However, this article will assume this is inaccurate.

8. Such as *Turritopsis dohrnii*, also known as the Immortal Jellyfish, which, when damaged, starving, or old, reverts into a polyp and then, from this polyp, buds off into a genetically identical jellyfish. A process which it could repeat indefinitely.
9. Such immortality may not necessarily take the form of continued biological existence. The prospect of the immortal soul and such an idea's central role in multiple world religions testify to immortality's popularity, as do techno-scientific means of avoiding death like cryopreservation, extreme biohacking, or mind uploads.
10. For an elaboration on the nature of immortality, see the 'A Note on Agelessness and Immortality' section.
11. Broderick D. Introduction I: Machines of Loving Grace (Let's Hope). In: Blackford R, Broderick D, eds. *Intelligence Unbound: The Future of Uploaded and Machine Minds*. Chichester: Wiley Blackwell; 2014:1–10.
12. Whether the identity of this artificial mind would be the same as that of the biological one or simply a copy that perceives itself as identical but numerically distinct is an interesting question but not important here.
13. Gibson R. An immortal ghost in the machine?. *AJOB Neuroscience* 2023;14(2):81–3.
14. Bostrom N, Yudkowsky E. The ethics of artificial intelligence. In: Yampolsky RV, ed. *Artificial Intelligence Safety and Security*. New York: Chapman and Hall/CRC; 2018:57–69.
15. See note 14, Bostrom, Yudkowsky 2018, at 82.
16. Olson SD. *The "Homeric Hymn to Aphrodite" and Related Texts: Text, Translation and Commentary*. Berlin/Boston: De Gruyter; 2012.
17. *Adventure Time*. Created by Pendleton Ward. Cartoon Network, 2010–2018.
18. For more on the degrees of immortality, see: Cave S. *Immortality: The Quest to Live Forever and How it Drives Civilization*. New York: Crown Publishers; 2012.
19. In short, within a closed system, the level of entropy, that is, disorder, either remains constant or increases, meaning that the totality of existence, including organized structures like buildings, stars, bodies, or computers, invariably moves from order to disorder—from function to dysfunction.
20. There is the potential here to explore aging's nature and what it means to get older (i.e. move through time, the transition from typical life phases, undergo specific biological functions, or travel from where one is born to when they die). Such an exploration, however, sits well beyond this article's scope. However, for more on this subject, see: García-Barranquero P, Albareda JL, Díaz-Cobacho G. Is ageing undesirable? An ethical analysis. *Journal of Medical Ethics* 2023. doi:10.1136/jme-2022-108823.
21. Parfit D. *Reasons and Persons*. Oxford: Clarendon Press; 1987.
22. Häyry M. There is a difference between selecting a deaf embryo and deafening a hearing child. *Journal of Medical Ethics* 2004;30:510–2; Fahmy MS. On the supposed moral harm of selecting for deafness. *Bioethics* 2011;25(3):128–36; Bennett R. The fallacy of the principle of procreative beneficence. *Bioethics* 2009;23(5):265–73; Cohen G. Intentional diminishment, the non-identity problem, and legal liability. *Hastings Law Journal* 2008;60(2):347–75.
23. This claim is based upon a refutation of mind–body dualism, which, while not universally accepted, this article assumes.
24. By comparison, one could alter the genes responsible for eye color before birth without fundamentally transforming the identity of that eventual person.
25. Additionally, one might ask whether conferring death where even the possibility of it did not exist previously makes one's life not worth living. While this article explores this to some degree in the 'Possible Harms of Immortality' section, a full exploration must wait for a later article.
26. Savulescu J. Procreative beneficence: Why we should select the best children. *Bioethics* 2001;15(5–6):413–26, at 415.
27. Further work on this approach to AGI (im)mortality could undoubtedly be needed, but this is best saved for a later piece of work.

28. Miller G. Death. *Edge*; 2007; available at <https://www.edge.org/response-detail/10352> (last accessed 7 June 2023).
29. Singer P. Famine, affluence, and morality. *Philosophy & Public Affairs* 1972;1(3):229–43.
30. Foot P. The problem of abortion and the doctrine of double effect. *Oxford Review* 1967;1:5–15.
31. Thomson JJ. A defense of abortion. *Philosophy & Public Affairs* 1971;1(1):47–66.
32. He writes, “There is only one really serious philosophical problem, and that is suicide.” Camus A. *The Myth of Sisyphus*. New York: Vintage Books; 1955, at 3.
33. This is not the same as saying death itself is bad, as some, like Epicurus, might take umbrage with this claim. Instead, between life and death, the former is preferable.
34. Frankl V. *The Doctor and the Soul*. Winston R, Wintston C, trans. New York: Alfred A. Knoff Inc; 1957, at 73.
35. Fischer JM, Mitchell-Yellin B. Immortality and boredom. *The Journal of Ethics* 2014;18(4):353–72.
36. Harris J. Intimations of immortality: The ethics and justice of life-extending therapies. *Current Legal Problems* 2002;55(1):65–95.
37. Williams B. The Makropulos case: Reflections on the tedium of immortality. In: Williams B, ed. *Problems for the Self*. Cambridge: Cambridge University Press; 1973.
38. Williams contrasts this with conditional desires, which are desires a person wants to satisfy, independent of their drive to continue living, such as food, shelter, water, heat, or medical care. One may wish for these things without also wanting to continue living.
39. Kagan S. *Death*. New Haven; Yale University Press; 2012.
40. Nussbaum M. *The Therapy of Desire*. Princeton; Princeton University Press; 1994.
41. Cave S. *Immortality*. New York: Crown Publishers; 2012.
42. Scheffler S, Kolodny K, eds. *Death and the Afterlife*. Oxford: Oxford University Press; 2013.
43. May T. *Death (The Art of Living)*. Stocksfield, UK: Acumen Publishing; 2009.
44. Nagel T. After you’ve gone. In: *The New York Review of Books*; 2014, at 61.
45. Braun KA, Ellis R, Loftus EF. Make my memory: How advertising can change our memories of the past. *Psychology and Marketing* 2001;19(1):1–23.
46. Belshaw C. Immortality, memory and imagination. *The Journal of Ethics* 2015;19(3/4):323–48, at 338.
47. Blumenfeld D. Living life over again. *Philosophy and Phenomenological Research* 2009;79(2):357–86; Perrett RW. Regarding immortality. *Religious Studies* 1986;22(2):219–33.
48. Black I. Novel beings and assisted nonexistence. *Cambridge Quarterly of Healthcare Ethics* 2021;30(3):543–55.
49. For more on the issue of human immortality and overpopulation, see: Farrant A. *Longevity and the Good Life*. London: Palgrave Macmillan; 2010; Cutas DE. Life extension, overpopulation and the right to life: Against lethal ethics. *Journal of Medical Ethics* 2008;34(9):e7; Kass LR. *Life, Liberty and the Defense of Dignity*. New York: Encounter Books; 2004.
50. For a fictional example of how difficult this could be, see *Torchwood: Miracle Day*.
51. United Nations. *Convention on the Prevention and Punishment of the Crime of Genocide*. (adopted 9th December 1948, entered into force 12th January 1951) 78 UNTS 278.
52. Dihal K. Enslaved minds: Artificial intelligence, slavery, and revolt. In: Cave S, ed. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press; 2020.
53. For example, see the eighteenth-century French Revolution.
54. While this is a staple of science fiction thinking, it must be noted that these narratives may depict a version of an AI takeover that is, to put it mildly, wildly inaccurate and distracts us from the immediate negative impacts AI might have on human well-being. See: Nature, stop talking about tomorrow’s AI doomsday when AI poses risks today. *Nature* 2023;618:885–6.
55. Bostrom N. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press; 2014, at 123.
56. Indeed, if one adheres to the precautionary principle, regardless of how slight the chance of a considerable risk is, measures to prevent serious harm must be taken wherever possible and, as what we are talking about here is the possibility of human enslavement, making AGI mortal rather than

immortal seems reasonable, especially so when such a position could be supported by other principled and pragmatic reasons, as has been suggested.

57. A point which should be raised, but sits outside the article's scope, is that a truly good AGI—even a morally perfect one—would also be limited by such aging. While mortality would ensure the death of an overlord, it would do so equally for an angelic AGI. Whether we are willing to pay this price needs further exploration.