AI & SOCIETY
https://doi.org/10.1007/s00146-024-01863-w

MAIN PAPER

Freedom, AI and God: why being dominated by a friendly super-AI might not be so bad

Morgan Luck[1]

Received: 8 October 2023 / Accepted: 5 January 2024

Abstract
One response to the existential threat posed by a super-intelligent AI is to design it to be friendly to us. Some have argued that even if this were possible, the resulting AI would treat us as we do our pets. Sparrow (AI & Soc. https://doi.org/10.1007/s00146-023-01698-x, 2023) argues that this would be a bad outcome, for such an AI would dominate us—resulting in our freedom being diminished (Pettit in Just freedom: A moral compass for a complex world. WW Norton & Company, 2014). In this paper, I consider whether this would be such a bad outcome.

Keywords AI · Friendly AI · Super-AI · Freedom · God · Domination · Pettit · Sparrow

1 Introduction

There is a concern that advances in AI may eventually lead to the development of a super-intelligent AI (or super-AI)—an AI that is far smarter than us. The concern is that such an AI might become so smart that, as AI researcher Paul Christiano remarked, "If, God forbid, they were trying to kill us, they would definitely kill us" (Adams and Hoffman 2023). Such concerns have led many of us to think we should develop AI in such a way that it would not try to kill us.

One such way is to develop *aligned* AI—AI that only has goals that appropriately align with our own. However, some have suggested that the best we can hope for is *friendly* AI—AI that is generally well-disposed to us (but may have some goals that do not align with our own). Such a friendly AI, as computer scientist Marvin Minsky famously remarked, might "If we're lucky…decide to keep us as pets" (Darrach 1970).

Some have embraced this best case scenario, with Apple co-founder Steve Wozniak stating "We want to be the family pet and be taken care of all the time" (Gibbs 2015). Others are less eager to roll over and let AI tickle our bellies—with philosopher Robert Sparrow (2023) recently arguing that such an outcome would result in a significant loss to our freedom.

A lot turns on who is right here—as whether it is prudent to develop super-AI is (at least partly) determined by the best case scenario. By way of analogy, consider whether or not it is a good idea to board a plane. The best case scenario is normally that the plane will arrive at its destination safe and sound, and the worst is it will crash, killing everyone. Knowing this, and the relevant odds, most of us still board planes. But imagine that, while the worst case scenario remains the same (everyone dies), the best case scenario is now that the plane is hijacked, and only after a traumatic few days does everyone safely disembark. Without even knowing the relevant odds, we can clearly see it would be a bad idea to board this plane.

If Minsky is correct, and the best case scenario is that we will be to AI as pets are to us, then whether or not we should board the AI plane will turn on whether or not such an outcome is good or bad. If it is good (as Wozniak suggests) then we still need to know the relevant odds (of this and the other possible outcomes) before making the decision to develop super-AI. But if this best case scenario is bad (as Sparrow argues), we can stop right there.[1] This is why Sparrow's analysis of this outcome is important.

My aim here is to evaluate this analysis. To do this I shall first outline Sparrow's argument (Sect. 2), then apply it to another super-intelligent agent—God (Sect. 3). The point of this move is to consider whether

[1] Assuming all the other options are not worse.

✉ Morgan Luck
moluck@csu.edu.au

[1] School of Social Work and Arts & Artificial Intelligence and Cyber Futures Institute, Charles Sturt University, Wagga Wagga, NSW 2650, Australia

Published online: 23 February 2024

Sparrow's argument extends too far. I shall then examine potential disanalogies that may limit such an extension (Sects. 3.1–3.4). Finally, I shall consider an argument for why the loss of freedom resulting from a super-AI (or God) existing may not actually be bad.

2 Sparrow's argument

Sparrow (2023) argues that even if a super-AI were to be friendly, it would still dominate us. This dominance would in turn diminish our freedom. Consequently, given we should want to be free, we should not want to develop a friendly super-AI. We can present this argument as follows,

1. We should want to be free.

2. If we should want to be free, then we should not want anything to dominate us.

3. If we should not want anything to dominate us then we should not want a friendly super-AI to exist.

So,

4. We should not want a friendly super-AI to exist.

Although it is premise 3 that I wish to focus on, it is worth quickly touching upon premises 1 and 2 first.

Sparrow draws upon Philip Pettit's republican theory of freedom (1997, 2012, 2014, 2016) to establish premises 1 and 2. Premise 1 (we should want to be free) seems uncontroversial given most of us hold freedom to be an important good. However, Pettit argues that freedom is also a necessary condition for other goods such as justice, democracy, and sovereignty (2014, pt. 2). So, if Pettit is correct, there is considerable reason to desire freedom.

The conception of freedom that Pettit champions is that of non-domination.

> Freedom requires that there be no controller or dominus, even one who gives you great latitude or leeway in your choices. In a word, freedom requires non-domination. (2014, 301)

Imagine a bridled horse that is given "free" reign by its rider to go where it pleases. Pettit argues that such a horse is not actually free, as its ability to go where it pleases is subject to the whim of the rider. One dominates the other '… to the extent that (1) they have the capacity to interfere (2) on an arbitrary basis (3) in certain choices that the other is in a position to make' (1997, 52). Given that freedom (in this sense) equates to non-domination, the truth of premise 2 seems to follow—if we should want to be free, we should not want to be dominated.

Sparrow supports premise 3 by arguing that the power a friendly super-AI would have over us, even if it never exerted this power, would be sufficient to constitute dominance.

> Insofar as it would remain true of such a machine that, if it wanted to "eat" us, it could, it seems that we would still be subject to its whims, dominated, and thus unfree. Domination exists, according to the republican tradition, where our rulers have the capacity to interfere arbitrarily, regardless of whether they are motivated to do so. (2023, 4)

Sparrows argument for 3 can be presented as follows,

5. If we should not want anything to dominate us then we should not want anything that has the capacity to interfere arbitrarily in our choices to exist.

6. A friendly super-AI would have the capacity to interfere arbitrarily in our choices.

So,

3. If we should not want anything to dominate us then we should not want a friendly super-AI to exist.

Most would agree that a super-AI would have the capacity to interfere in our choices (Bostrom 2014; Yudkowsky 2008). Where things get interesting is determining whether a friendly super-AI could do this arbitrarily.

Pettit takes someone to have the capacity to arbitrarily interfere in our choices if they "have the ability to interfere intentionally in one of the options without your permission or control" (2016, 51). Arnold and Harris (2017), put it like this,

> B's power over A is arbitrary insofar as it is not reliably constrained by effective rules, procedures, etc., that give A control over B's exercise of that power. (58)

In other words, if there is no reliable way for a person to exert control over the power that is interfering in their choices then this interference is arbitrary. Pettit refers to this as uncontrolled interference: "interference that is uncontrolled by the person on the receiving end" (Pettit 2012, 58). For example, if our bridled horse has no control over the power the rider might use to interfere with their choices, such interference would be arbitrary.

So, given such an understanding, would a friendly super-AI arbitrarily interfere with our choices? Or would the super-AI's friendliness curtail its ability to arbitrarily interfere? To help us answer such questions let us, perhaps unexpectedly, look to God.

3 What's God got to do with it?

Sparrow provides us with reason to think that even if a super-AI were friendly, it would still dominate us; so, perhaps we should not want such an AI to exist. But what if we were to make a similar argument regarding God? Should we, by the same light, want God not to exist? And if not, why not?

Now at this juncture you might think “Look it’s hard enough trying to think about a super-AI—why introduce a more ineffable agent into the mix?” There are a few reasons. First, it is difficult to think clearly about agents with properties (such as, intelligence or power) that far exceed our own. However, philosophers of religion have a head start here, as they have been trying to do exactly this for well-over a millennium (take, for example, Augustine of Hippo’s [354–430 C.E.] analysis of God in *The Trinity* 1990). So, building upon this work, a comparison with God may help us think more clearly about super-AI.

Second, Pettit himself, when considering whether humans could establish a republican state free from domination, raises the possibility that only an AI or God might have what it takes to do so.

> I have tried to identify the things that a republican state should try to do in order to counter the domination that can go with private dominium. And I have attempted to spell out the measures that it should adopt in order to reduce the presence of arbitrary will in its own coercive arrangements and so to guard against domination by state imperium. But these recommendations, for all I have said, may only have the character of an unattainable wish-list. Perhaps only creatures of an unattainable cast of mind—god-like creatures, machine-like creatures, or whatever—could sustain the republican regime described. (1997, 206)

Pettit's reference to god-like and machine-like creatures may merely be a rhetorical device used to underline the difficulties associated with power being held by humans. However, the device only works because such agents (being both incorruptible, super-intelligent, etc.) are more promising candidates for establishing a free state. So, Pettit’s own intuition on this matter may run counter to Sparrow’s; not only regarding the suitability of super-AI to help us establish a free state, but perhaps also, if Sparrow’s argument generalises, the suitability of God to do the same.

Third, there are some interesting cognitive connections between super-AI and God. Spatola and Urbanska (2020) report that, when thinking about “artificial intelligence and robots, people appear to draw parallels to divine entities” (329). And when people are asked to think about AI in relation to God, as Karataş and Cutright (2023) report, they are more likely to think positively about AI. I hope to draw upon this positive parallel to put pressure on Sparrow's concern that super-AI might dominate us. My aim is to suggest that if you are a pro-theist (i.e. someone who wants God to exist—even if you do not think they do), then one must, given the validity of Sparrow’s argument, reject the truth of at least one of its premises.

Sparrow’s main argument can be repurposed as an anti-theistic argument as follows,

1. We should want to be free.

2. If we should want to be free, then we should not want anything to dominate us.

3. If we should not want anything to dominate us then we should not want God to exist.

So,

4. We should not want God to exist.

If you are convinced by Sparrow’s argument, and God is relevantly similar to a super-AI, then it seems one should also be an anti-theist (someone who does not want God to exist—even if you believe they do). This may be a surprising result for those that champion Petit’s republican sense of freedom. So, let us see what might be done to resist it.

Let us begin by focusing our attention on premise 3. Consider an analogous argument supporting 3.

5. If we should not want anything to dominate us then we should not want anything that has the capacity to interfere arbitrarily in our choices to exist.

6. God would have the capacity to interfere arbitrarily in our choices.

So,

3. If we should not want anything to dominate us then we should not want God to exist.

But is 6 true? Would God have the capacity to interfere arbitrarily in our choices? Perhaps this is a relevant difference between a super-AI and God. A difference that might explain why Sparrow’s argument against AI is sound, but not the modified argument against God.

Whether or not God has the capacity to interfere arbitrarily in our choices will depend on our conception of God. The Old Testament God that, for example, played a role in the death of Job’s children might seem like an agent that has such a capacity. For this, interference was certainly “uncontrolled by the person on the receiving end” (Pettit 2012, 58). In which case, 6 would be true. So, rather than adopting a theological account of God informed by a particular religious tradition, what if we instead considered a more philosophical (and minimal) conception of the divine?

God, as conceived by philosopher Richard Swinburne, is an agent who is omnipotent, omniscient, and perfectly free (2004). From these primary properties, Swinburne argues various secondary properties follow. For example, “God’s perfect goodness follows deductively from his omniscience and his perfect freedom” (99). This is because, put briefly, an omniscient agent would know what actions are morally best, and a perfectly free agent (one not subject to non-rational influences) would “always do any action that he believes to be the best action available to him” (104).

Although this conception of God suits our purposes, it is worth noting that alternative notions may also fit the bill. All that is required is for God to possess at least as much power, knowledge and goodness as our friendly super-AI. We need not commit ourselves to which of these properties are primary and secondary, or what the limits of such properties may be.[2] Such a minimally committed conception even leaves open the possibility that God never has, and perhaps never will, intervene in the natural world (i.e. a deistic God). All we require is that God could intervene if they so choose. Such a conception of God may well be compatible with some religious traditions, but it assumes less.

So, could such a God arbitrarily interfere in our choices? To find out, let us examine Sparrow's reasons for why a friendly super-AI could arbitrarily interfere to determine if they also apply to God.

3.1 Perfectly good interference

Sparrow first considers whether a friendly super-AI, given its benevolent disposition towards us, would only interfere when it is in our interests. Sparrow determines that such a possibility does not protect us from domination.

> …a benevolent dictator, who only interferes in our lives when it is in our interests, is still a dictator and his/her power is still inimical to our freedom. (2023, 3)

So, something being benevolent does not rule out the possibility of it arbitrarily interfering.

Pettit makes a similar point regarding paternalism. He argues that an agent is paternalistic when they interfere in your choices according to what they interpret to be your best interests (although this may not align with what you think is in your interests).

> And to the extent that I impose my own interpretation on your interests, discounting yours as inferior, I act paternalistically…Such paternalistic intervention, in the nature of the case, involves interfering according to my own arbitrium, or ‘will’, not yours, and is an exemplar of domination. (2012, 59)

Of course, a super-AI may well have a superior interpretation of what was in our interests. But the point here is that, regardless of whether the interference was benevolently motivated or in our best interests, if we have no control over it, it will constitute as arbitrary (or uncontrolled) interference. These points, however, seem to apply equally well to God.

An omniscient agent will have a superior interpretation of what is in our best interests. And any interference from a perfectly good agent will be benevolently motivated (or at least not be malevolently motivated). But again, if we do not have the right kind of control over this inference, then, according to Sparrow, it will constitute an arbitrary (or uncontrolled) interference. So God's goodness *alone*, like the super-AI’s friendliness, does not seem to be an obstacle to domination.

3.2 The ultimate check of power

Sparrow next considers whether the friendliness of a super-AI might compel it to avoid arbitrary interference by only interfering in our choices after granting us an appropriate amount of control—by for example “asking us what we want and listening to our deliberations” (2023, 3). Following Russell (2019), we are asked to imagine the possibility of working with the super-AI to help us achieve our own ends in a manner that does not obviously jeopardise our freedom.

In a similar fashion, God’s goodness might also guarantee that humans are granted an appropriate amount of control.[3] In which case, before intervening in a manner that would affect our choices (by, for example, flooding the world) God would properly consider our preferences. Yet even if this were the case, Sparrow raises an additional problem.

There is a tension, Sparrow claims, between “the intelligence of AI, its power, its freedom, and our freedom” (2023, 4). The super-AI, he argues, even if it worked with humans to achieve our ends, would still dominate us. To make this point Sparrow draws our attention to the relationship between a government and its citizens.

> In order for the power of the state to be compatible with the liberty of citizens, the government must listen to reason and justify its exercise of power in terms that citizens accept. However, it must also be hostage to reason in the sense that if the citizenry is not convinced that the government’s exercise of power is justified, the government’s power is checked. The ultimate check on the power of governments is the capacity of the citizenry to overthrow them. (4)

So, although a government may work with its citizens to achieve their ends, the type of control citizens ultimately require in order not to be dominated is the ability to overthrow their governments.

However, the intelligence, power and freedom of the super-AI is such that Sparrow doubts our ability to overthrow it. In which case, despite its friendliness potentially

[2] The exact nature of such divine powers are the subject of ongoing debate. A debate that raises such questions as: could God, given their omnipotence, make the impossible possible (e.g. the paradox of the stone); do we really have free will if God’s omniscience fixes future events; and, is God’s perfect goodness compatible with avoidable evils (see Hoffman & Rosenkrantz 2008).

[3] Note that some religious traditions might hold that a measure of control has already been granted by God in the form of free will. However, free will is only a necessary condition for Pettit's sense of control.

compelling it to work with us, and for us (as governments should), the AI would still be dominating us.

It is not hard to see that God would be dominating us in a similar way. God's intelligence, power and freedom surpasses that of a super-AI. So, even if God were to work with us to achieve our own ends, given our inability to remove God from its position of power, God would still, following Sparrow's line, dominate.

3.3 A conceptual limitation on agency

Sparrow also makes a minor conceptual point for why we could not "hardwire" a friendly super-AI to not act against our interests. This is because such a hardwired AI would not technically be a super-AI—a type of super-intelligent *agent*.

> It is unclear whether the existence of such hardwired limits on what an AI is capable of desiring is compatible with claiming it to be an agent and, therefore, "genuinely" intelligent (2023, 4).

The thought here is that if something cannot hold certain desires, then that thing would not be an agent, in which case it cannot be truly intelligent (let alone super-intelligent). So, whatever it is that may be taking away our freedom, it would not be a genuine super-AI.

This point has merit; however, limits need to be established to avoid overgeneralization. For example, you are (hopefully) presently incapable of forming a genuine desire to torture an innocent child just for fun. Although you cannot form this desire (perhaps due to some biological and/or psychological "hardwiring"), this does not suggest you are not an agent. Of course, this is just a single desire—presumably genuine agency can tolerate such a limitation. Sparrow may instead have in mind a much larger set of desires that a super-AI could not hold because of its hardwired friendliness. A set that may be large enough for us to rightly question its agency.

Yet a similar problem seems to extend to God. God's nature is such that they will be unable to hold numerous desires. For example, being perfectly good (something they must be by definition) they will be unable to hold any immoral desires. Likewise, being perfectly free they will be unable to hold any irrational desires. Yet, despite God not having the ability to be otherwise, such limitations are typically not thought to be an obstacle to their agency.[4] So, if God's agency is able to withstand their inability to desire to act in an immoral or irrational way, the possibility arises that a super-AI's agency may be able to withstand their inability to desire to act in an unfriendly way.

3.4 Eyeballing God

Sparrow presents Pettit's "eyeball test" (1997, 71–73) as further reason to think a friendly super-AI would dominate us.

> In a republic, citizens meet as equals of a certain sort. Even if some are wealthy and some are poor, no citizen dominates another. Knowing that they are secure from the arbitrary exercise of power by others, citizens do not need to bow and scrape to their "superiors". They can look each other in the eye. (4)

Put simply, if A and B can look each other in the eye (that is, treat each other as equals) then A is not dominating B and vice versa.

Sparrow points out that we will be unable to look a super-AI in their metaphorical eyeballs. This is because, "we cannot have a relationship of equals with a superintelligence, because we will not be its equals. The power that even a friendly superintelligence would have over us means that we would effectively still be its pets"(4). But, predictably, this point also applies to God.

The power and intelligence of God far exceeds that of a super-AI. In which case, our relation to God would also fail the eyeball test. In which case, Sparrow has provided another reason to think God's existence is incompatible with our freedom.

4 Optimal freedom

In the previous Sects. (3.1–3.4), we examined Sparrow's reasons for why, if a friendly super-AI existed, it would dominate us. This was to determine if the reasons supporting premise 3 of Sparrow's argument might also apply to God. A pro-theist who is convinced by Sparrow's reasoning might have hoped that we would find a relevant disanalogy between a friendly super-AI and God, such that none of Sparrow's reasons would hold in respect to God. However, the opposite seems to be the case—they all seem to. Nevertheless, there is another avenue open to the pro-theist. They might instead reject premise 1—that we should want to be free.

Sparrow holds that what "would be required to preserve human freedom is that a friendly AI could not act against humanity's interests" (4). But, following Bostrom (2014) and Yudkowsky (2008), he is doubtful that we could develop such a super-AI. Consequently, a super-AI, even a friendly one, would always be dominating us. But should we really want a friendly super-AI that could never act against humanity's interests?

Consider the following possible future,

> Humanity develops a friendly super-AI that *can't act against our interests*. Using the super-AI we create tech-

[4] Or their omnipotence—see Swinburne (2016, ch.9).

> nologies that transform humanity forever. But (despite the super-AI's efforts) these technologies facilitate extreme decadence. Slowly corrupted by our strange new transhuman desires we turn our eyes fearfully to the rest of the galaxy. To secure our lives of endless luxury we decide to preemptively strike out at any alien civilization that either has, or is close to having, super-AI. In time humanity destroys or dominates all sentient life everywhere.

Compare this to an alternative future,

> Humanity develops a friendly super-AI, but one that *can act against our interests if appropriate*. Using the super-AI we create technologies that transform humanity forever. But (despite the super-AI's efforts) these technologies facilitate extreme decadence. Slowly corrupted by our strange new transhuman desires we turn our eyes fearfully to the rest of the galaxy. To secure our lives of endless luxury we decide to preemptively strike out at any alien civilization that either has, or is close to having, super-AI. However, our super-AI steps in and stops us before we can enact our plan.

Despite the second future containing a super-AI that dominates us, that is the better future, that is the better AI. Admittedly, it is not better for humanity in respect to their own interests at that time. But it is better all things considered. So, what does this reveal?

We should not want a super-AI that never acts against our interests. Instead we should want a super-AI that never acts against our interests for the wrong reasons. This provides our pro-theist with a potential out—for the same is true of God. Perhaps we should not want to be completely free. In other words, perhaps premise 1 is false.

If the only way to stop some group (such as a terrorist organisation, or rogue nation) from causing some moral atrocity is by occasionally, or even systematically, interfering in their choices in ways that they cannot control, then arguably they should be dominated. However, in doing so their freedom is diminished. But perhaps this is not such a bad result when what is dominating is properly orientated (as God should be). Such a dominator would be orientated to give us as much freedom as possible, but not so much that it could not interfere if it was the right thing to do. That is, it would seek to optimise our freedom.

It should be noted that it may sometimes be hard to know if interfering is the right thing to do. For example, it seems clearly wrong to stop a parent scolding their child in a slightly more harsh manner than was called for. Such interference does not seem worth compromising the parent's freedom to raise their child. But, it seems clearly right to stop them burning their child with a cigarette as a form of punishment. Here the interference seems worth the compromise. And presumably there will be borderline cases where things are not so clear. Of course, given their intelligence, the borderline will be far crisper for a super-AI or God. But the point here is that, merely preventing unjust harm does not make interfering necessarily the right thing to do. The value of such interference must be weighed against the cost of compromising our freedom.

If this is correct, then perhaps premise 1 should be modified to something like this,

> We should want to be free, but not to such an extent that we cannot be dominated for the right reasons.

Such a modification suggests we should not want to be maximally free. Instead we should only want to be optimally free—free as much as we can give the possible existence of a dominator that would only interfere if it was right to do so.[5] At first, this might seem like an extraordinary curtailing of our freedom. But it has some support from a very ordinary practice—parenting.

Anca Gheaus (2021) argues that parents should be able to dominate their children. This is because, the removal of the mere capacity of parents to arbitrarily interfere in their children's choices would be worse than limiting their children's freedoms.

> Child-rearing without domination would require the elimination of the possibility to use, with impunity, power over children in ways that do not track their interests—a goal that is unattainable without the sacrifice of other, more important (non-republican) goals: children's general interest in adequate care, including their shared interest in intimacy. (756)

Gheaus does not think parents should be able to do whatever they want. But rather parents should uphold the "least dominating child-rearing arrangement" (749) they can. In other words, we should want children to be as free as they can given that their parents do, and should, dominate them. A result which mirrors the modified premise 1.

This is not to suggest that there are not important differences between children and parents, and humans and God/super-AI.[6] For example, the mental capabilities of children make them vulnerable in ways that adults are not, such that the benevolent domination of children may be permissible (or even required)—whereas this is not the case for adults.

[5] Some may object to the notion of a reduced amount of freedom, on the basis that freedom is absolute—and so does not come in degrees. However, Pettit's (1997) sense of freedom is not absolute, as "non-domination comes in degrees both of intensity and extent" (273).

[6] Although there are also important similarities in the asymmetries of power, intelligence and freedom that these parties hold in respect to each other. And if one held that these groups were relevantly similar then an argument could be made that humans, like children, are not mature enough to self-govern. A result that would permit (or even require) God, or a friendly AI, to dominate. I do not make such an argument here.

Rather, the point of the parental example (as with the previous example of our transhuman future) is to suggest that dominance is sometimes permissible—it is only when we are wrongly dominated that we should have an issue.[7]

Recognizing this may take some of the sting out of the tail of Sparrow's concern that the existence of a super-AI may reduce our freedom. Perhaps we also should not desire maximal freedom. Not because we are vulnerable in the same way children are, but because we want to allow for the possibility of being rightly dominated. And if we should not desire maximal freedom, then this may provide us with the opening we need to possibly want a God, or a friendly super-AI, to exist.

5 Concluding remarks

I am not claiming that the development of a friendly super-AI would be good. It may well be very bad. Rather I am claiming, if it is bad, it would not be because it would result in us having less freedom—providing our freedom is reduced by the right kind of agent. That is, an agent that would seek to optimise our freedom—to give us as much freedom as possible, but not so much it could not interfere for the right reasons.

Although this result may be particularly attractive for a pro-theist (as it allows them to resist the conclusion that we should not want God to exist), it is worth quickly noting that the same move could be made in regard to any suitably intelligent agent. For example, an alien civilization that is so technologically advanced it could interfere in our choices if they wanted to, but also so incredibly benevolent and intelligent that they would not, unless this was the right thing to do. Should we want there to be no such benevolent civilizations out there? Such a desire seems suspect (it strikes me as a particularly egregious instance of anthropocentricity).

The real question we should be focusing on here is whether a friendly super-AI would be the right kind of agent to permissibly dominate us. One that would try to optimise our freedom. I do not feel particularly confident that it would be—but I hopefully would not begrudge its existence if it were, despite the loss of freedom that might result.[8]

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data availability Data sharing is not applicable, as this paper does not analyse or generate any datasets.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Adams R, Hoffman D (Hosts) (2023). 'How we prevent the AI's from killing us with Paul Christiano' [Video podcast episode]. Bankless. (Apr 24) https://www.youtube.com/watch?v=GyFkWb903aU. Accessed 3 Dec 2023

Arnold S, Harris JR (2017) What is arbitrary power? J Polit Power 10(1):55–70

Augustine, of Hippo, Saint, 354–430 (1990) The Trinity. Hill, E., Rotelle, J. E. (Trans) United Kingdom: New City Press

Bostrom N (2014) Superintelligence: paths, dangers, strategies. Oxford University Press, Oxford

Darrach B (1970) Meet Shaky the first electronic person: the fascinating and fearsome reality of a machine with a mind of its own. In: Life Magazine (Nov, 20). Time Inc., New York

Gheaus A (2021) Child-rearing with minimal domination: a republican account. Political Studies 69(3):748–766

Gibbs S (2015) Apple co-founder Steve Wozniak says humans will be robots' pets. The Guardian. (June, 25) https://www.theguardian.com/technology/2015/jun/25/apple-co-founder-steve-wozniak-says-humans-will-be-robots-pets. Accessed 3 Dec 2023

Hoffman J, Rosenkrantz GS (2008) The divine attributes. Wiley, Berlin

[7] This point might be recast in terms of paternalism vs. liberalism; that we should be prepared to accept some degree of paternalism provided it is worth the reduction in liberty that may result. Some take the permissible degree of paternalism to be quite small (as exemplified by libertarian paternalism [Thaler and Sunstein 2012]—where the use of small nudges are used to influence people's choices), whereas others permit a larger degree (such a Kleinig's argument from personal integrity [1983, 68]—which allows more direct interference providing it respects others ranking of their own interests/projects). In this paper, I suggest that whatever degree of paternalism is warranted is appropriate to be desired.

[8] My thanks to Michael Shepanski for his help on this paper.

Karataş M, Cutright KM (2023) Thinking about God increases acceptance of artificial intelligence in decision-making. Proc Natl Acad Sci 120(33):e2218961120

Kleinig J (1983) Paternalism. Manchester University Press, Manchester

Pettit P (1997) Republicanism: a theory of freedom and government. Oxford University Press, Oxford

Pettit P (2012) On the people's terms: a republican theory and model of democracy. Cambridge University Press

Pettit P (2014) Just freedom: a moral compass for a complex world. WW Norton & Company

Pettit P (2016) The globalized republican ideal. Glob Just Theory Practi Rhetoric 9(1):51

Russell S (2019) Human compatible: AI and the problem of control. Allen Lane, Bristol

Sparrow R (2023) Friendly AI will still be our master. Or, why we should not want to be the pets of super-intelligent computers. AI & Soc. https://doi.org/10.1007/s00146-023-01698-x

Spatola N, Urbanska K (2020) God-like robots: the semantic overlap between representation of divine and artificial entities. AI & Soc 35(2):329–341

Swinburne R (2004) The existence of god. Oxford University Press, Oxford

Swinburne R (2016) The coherence of theism. Oxford University Press, Oxford

Thaler RH, Sunstein CR (2012) Nudge: the final edition. Penguin Books Limited, United Kingdom

Yudkowsky E (2008) Artificial intelligence as a positive and negative factor in global risk. In: Bostrom N, Cirkovic MM (eds) Global catastrophic risks. Oxford University Press, Oxford, pp 308–345