Journal of Ethics and Emerging Technologies

IEET

Article

On Artificial Superintelligence and the Problem of Charismatic Extinction Threats

Nicholas Agar [1]* and Murilo Vilaça [2]

1 University of Waikato; nicholas.agar@waikato.ac.nz
2 Fundação Oswaldo Cruz; murilo.vilaca@fiocruz.br
* Correspondence: nicholas.agar@waikato.ac.nz

Abstract: This paper focuses on the challenge of finding a rational response to Artificial Superintelligence (ASI) as an extinction threat. We allow that artificially superintelligent beings are possible. This leaves open the question of how much we should worry about them. We treat ASI as a *charismatic* extinction threat. Our starting analysis of charisma comes from the sociologist Max Weber (1947). We extend the concept of charisma beyond individual personalities to events including extinction threats. Our principal example of a charismatic extinction threat is Skynet, the human-unfriendly AI of the movies of the *Terminator* franchise. Skynet's charisma interferes with the processes by which we rationally evaluate future risks. Our exploration of the psychological and emotional dimensions of assessing extinction threats considers work by the Nobel laureate economist Robert Shiller (2019) in the emerging field of narrative economics. We connect the virality of extinction stories with the work of the psychologist Elke Weber. According to Elke Weber (2010) we have a *finite pool of worry* to allocate to all of our future concerns. The charisma of Skynet means that we risk worrying too much about it and, as a consequence, worrying insufficiently about the uncharismatic challenge of climate change. We conclude with a brief discussion of a proposal that could lead to a more rational allocation of worry about extinction and other threats to humanity. We counsel imagination insurance for an intrinsically uncertain future.

Keywords: Artificial Superintelligence; Extinction Threat; Charismatic Extinction Threat; Finite Pool of Worry

Citation: Agar, Nicholas and Vilaça, Murilo. 2025. On Artificial Superintelligence and the Problem of Charismatic Extinction Threats. *Journal of Ethics and Emerging Technologies* 35: 2. https://doi.org/10.55613/jeet.v35i2.166

Received: 21/02/2025
Accepted: 21/02/2025
Published: 10/03/2025

Publisher's Note: IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The arrival of ChatGPT triggered a great deal of speculation about artificial superintelligence. In this paper, we define an Artificial Superintelligence (ASI) as a form of artificial intelligence that greatly surpasses human intelligence in all aspects – cognitive, emotional, and practical.

Discussion of the dangers of ASI has occurred across the academy. In 2014, the philosopher Nick Bostrom introduced the control problem – "the problem of how to control what the superintelligence would do (Bostrom, 2014, p. iv). How can we form a rational plan to control an entity whose intelligence is many magnitudes more powerful than ours?

This leads Bostrom to propose that "existential catastrophe" is "a plausible default outcome" of progress in AI (Bostrom, 2014, p. 115). Bostrom describes an intelligence explosion that predictably takes an AI from the human-level capacities of an Artificial General Intelligence (AGI) to an Artificial Superintelligence. The physicist Max Tegmark (2017) echoes Bostrom's concerns. He explores "fast takeoff" scenarios in AI, in which it takes a matter of days, not decades, for a single entity to control Earth. The computer scientist Roman Yampolskiy (2015) has explored various ways in which we might seek to control an emerging artificial superintelligence. Yampolskiy (2015, acknowledgments) cites his children—the "only two human-level intelligences I was able to create thus far" – as reasons to doubt that we can "fully control any intelligent agent." For Yampolskiy, the potential creation of ASI ups the ante. If he struggles to control his children, what are the consequences of failing to control an ASI?

We offer a response to artificially superintelligent beings that concedes they are possible and potentially imminent. An ASI violates no law of logic or of physics that we know of. Suppose that an artificial superintelligence arrives and is human-unfriendly – behaving as the supercomputer Skynet in the Terminator movie franchise does, seeking to send humanity extinct. That would be very bad, for us at least. Such an AI would need to be vastly different from the Large Language Model AIs that are currently writing student essays. But the shock at the capacities and abrupt arrival of ChatGPT suggests that advances in AI that could create an ASI should not be airily dismissed.

This article treats ASI as a *charismatic* extinction threat. Our starting analysis of charisma comes from the sociologist Max Weber (1947). Weber explains how he understands the concept – "The term 'charisma' will be applied to a certain quality of an individual personality by virtue of which he is set apart from ordinary men and treated as endowed with supernatural, superhuman, or at least specifically exceptional powers or qualities" (Weber, 1947, p. 358).

We extend the concept of charisma beyond individual personalities to events including extinction threats. Our principal example of a charismatic extinction threat is Skynet, the human-unfriendly AI of the movies of the *Terminator* franchise. It tends to elicit a different response from that prompted by the uncharismatic extinction threat of climate change.

We focus on the propensity for Skynet's charisma to interfere with the processes by which we rationally evaluate future risks. Our exploration of the psychological and emotional dimensions of assessing extinction threats considers work by the Nobel laureate economist Robert Shiller (2019) in the emerging field of narrative economics. According to Shiller stories can go viral, influencing our economic choices for good or ill. The same phenomenon applies to our assessment of extinction threats. We propose that Hollywood movies can be effective vehicles for the virality of stories about human extinction. Often Hollywood virality can cause us to worry too much about a potential cause of extinction.

We connect the virality of extinction stories with the work of the psychologist Elke Weber. According to Elke Weber (2010) we have a *finite pool of worry* to allocate to all of

our future concerns. The charisma of Skynet means that we risk worrying too much about it and, as a consequence, worrying insufficiently about the uncharismatic challenge of climate change.

We conclude with a brief discussion of a proposal that could lead to a more rational allocation of worry about extinction and other threats to humanity. We counsel imagination insurance for an intrinsically uncertain future. If we are to challenge an approach overly focused on the threat from Skynet, we need to better leverage human diversity to better hear stories about the dangers and opportunities of the future that Hollywood has thus far not seen fit to feature in any blockbuster movie franchise.

2. Responding to the Unknown

The perception of the risk of something happening or the risk associated with something can vary depending on a variety of factors (Vilaça & Lavazza, 2022). A traumatic experience with something (for example, an illness in the family or a widely publicised plane crash) can lead us to systematically overestimate the danger from these negative events. Conversely, good experiences can bias us towards overly optimistic assessments of future possibilities.

We seek a framework for rationally responding to future risks, some of which we are aware of and others of which we are currently oblivious. A rational assessment of the potential dangers and benefits from future activities demands responses to distortions in the ways we think about them.

One factor that can interfere with our perception of the threat posed by something is simple unfamiliarity. It's difficult to take a rational stance on something we don't know about. The unknown can have a mobilizing force, which can be expressed ambiguously, that is both attractively and repulsively. The degree of our ignorance about a future event can sometimes inspire hope. In other cases it exacerbates fear. If Max Weber is right charismatic things enjoy a special status, seemingly treated as if they have magical powers. That magic can potentially be used for good or ill.

Weber (1947) applies the concept of charisma to individual personalities. We extend his concept to extinction threats. The concept of charisma has also been applied to discussions about preserving biodiversity in the form of the phenomenon of charismatic species. Here too it can have a distorting effect on the rational allocation of moral concern.

Charismatic species attract significant interest and empathy from the public (Courchanp et al. 2018). We care a great deal about protecting animals like the Siberian tiger (*Panthera tigris altaica*). We care less about protecting the uncharismatic endangered sandy blind mole-rat (*Spalax arenarius*). The concept of charisma used by Courchanp et al. is more colloquial and they make no explicit reference to Max Weber. But we think is useful to consider connections between ecological charisma and Weber's usage tied to authority and leadership. Charismatic Siberian tigers command attention from us in a way that uncharismatic mole-rats do not. Suppose we sought to adjust our environmental priorities by giving the blind mole-rat more love so as to maximize the health of the biosphere. To make more room for *Spalax arenarius*, we may have to care less about charismatic endangered animals like *Panthera tigris altaica*.

We think that ASI has been elevated to the status of a charismatic extinction threat: something very powerful, extraordinary, superhuman, almost supernatural that can destroy us. For a philosophical rebuttal of this reasoning see Agar (2016). In fictional presentations ASI is highly threatening and uncontrollable. Because its thought processes are by definition beyond human comprehension we find ourselves unable to muster a rational response. A chess grandmaster predictably defeats a novice player by deploying stratagems beyond the novice's understanding. The novice can hope to acquire the understanding of the grandmaster. But the gap between human understanding and that of the ASI suggests that we cannot hope to replicate the intellectual achievement of the novice who studies hard and eventually develops into a world class chess player.

Apotheotic and apocalyptic depictions can be viewed as binaries in the presentations of charismatic future events. Will this choice lead to rapture or mass extinction? Will ASI end humanity or merely, as OpenAI founder Sam Altman sometimes speculates, end capitalism introducing a new more just way of sharing wealth?

3. Humanity's Large but Finite Pool of Worry

The fact that a scenario is both terrifying and compatible with the laws of physics and logic leaves open the question of how much we should worry about it.

Consider cancer. Cancer is one among many diseases that can kill humans. But it seems to have a special status among diseases. When placed alongside other killers of humans cancer seems to have Weberian charisma. It is "set apart from ordinary" diseases and "treated as endowed with supernatural, superhuman, or at least specifically exceptional powers or qualities". The title of the 2010 best seller on cancer written by the physician Siddhartha Mukherjee, "The Emperor of All Maladies," picks up on this charisma (Mukherjee, 2010). Mukherjee brings cancer closer to Weber's original account by calling his book "a biography". He gives cancer a personality.

Terminal cancer is a very bad, logically and physically possible outcome for each of us. But it is possible to be too worried about it. An ostensibly healthy person who resolves never to walk outside out of fear that even the briefest exposure to ultraviolet light could cause DNA damage, triggering a lethal cancer, is probably too worried. They may need reassurance that points to all the pleasures their excessive level of worry about cancer denies them. They should also think about the many less charismatic killers that leave humans just as dead as terminal cancer does.

The healthy cancer obsessive needs to relax and get out more. The fact that a human-unfriendly ASI is possible and that it could be very bad for us if it came into existence does not, by itself, suggest that we should worry much about it. There are many terrible futures consistent with the laws of logic and physics, as we currently understand them, that we shouldn't worry much about. The abrupt formation of a black hole at the centre of the Earth is not compatible with the laws of physics as we currently understand them but it is a logical possibility.

If you find a group of friends and give yourself an hour to brainstorm causes of extinction, you can almost certainly think up many horrible ways for humanity to die out that don't deserve much worry. The speculations of Bostrom (2014) about the creation of

an ASI lean heavily on the logical possibility of its creation. We can say that its creation is not incompatible with the laws of physics we *currently* understand them. Bostrom's book contains a host of speculations about the capacities of an ASI. Perhaps when humans or AGIs get closer to building one, they will discover a range of inconsistencies with the laws of physics that we today are oblivious of. Perhaps artificial superintelligence will join the list of logically possible causes of extinction that people worry less about once they have a better understanding of the physics. Halley's Comet has long been viewed as a harbinger of some significant event by those who witnessed it. In 1910 there were fears about what might happen to humans once the Earth passed through the Comet's tail (Kean, 2025). By the time the Comet returned in 1986 we heard less about this particular fear.

We frame the question of how much we should worry about an ASI in the terms offered by the psychologist Elke Weber. Weber is interested in our collective response to climate change. She describes us as drawing on a "finite pool of worry." Elke Weber treats worry as fungible, able to be transferred from one issue of concern to another. She says, "Unlike money or other material resources, which can be saved or borrowed, the amount of attention available to anyone to process the vast amount of information potentially available on innumerable topics is small and very finite" (Weber, 2010, p. 335). As we observe the worsening climate crisis, we find ourselves concluding that we did not grant climate change a sufficient claim on our finite pool of worry (Lynas et al., 2021). We should bemoan climate change's lack of charisma.

Elke Weber's framework applies to individuals reflecting on the climate crisis. We can ask how much of an *individual's* finite pool of worry should be allocated to climate change. But we can also ask about the problem's claim on our *collective* pool of worry. As we write these words in 2025, the global population stands at 8.1 billion. The collective pool of worry – the sum of all individual pools of worry – remains finite. But it grants humanity considerable reach to worry about our and the planet's problems. Individuals engaged in other particularly time-consuming and important pursuits – say, finding a significantly improved cancer treatment – might absolve themselves from any obligation to worry about climate change. But they should expect that others are sufficiently worried to compensate for their indifference. This is one of the collective benefits of the division of cognitive labour, according to which our patterns of deference to others can compensate for gaps in our understanding (Keil, 2006).

An implication of Elke Weber's framing of climate change is that, unless we can expand the collective pool of worry by worrying more, adding a new worry or worrying more about an established concern should lead us to reduce the claims of other concerns on our finite pool of worry. It's a problem that climate change could send humanity extinct. It's also a problem for our collective response that climate change is not an especially charismatic extinction threat.

4. Intrinsic and Extrinsic Causes of Worrying Less or More about an Extinction Threat

When asking about changes in the claim of a potential cause of extinction on our pool of worry, we should distinguish between intrinsic and extrinsic causes of change.

Intrinsic cause of change in how much an individual or a collective worries about a potential cause of extinction: The change occurs due to receipt of fresh evidence about the potential cause of extinction.

Extrinsic cause of change in how much an individual or a collective worries about a potential cause of extinction: The change does not occur due to the receipt of fresh evidence about the potential cause of extinction.

Suppose the latest Intergovernmental Panel on Climate Change (IPCC) report points to new scientific evidence that the damage caused by climate change is increasing more quickly than previously thought. That would be an intrinsic cause for increased worry. Suppose that the evidence justifying increased worry is later disconfirmed. That would be an intrinsic cause for decreased worry.

An extrinsic cause for changing our quantity of concern about a given threat offers no evidential support for that change. Suppose a new movie with high production values and winsome actors causes people to worry more about the climate crisis. That movie may increase the amount we worry about climate change without offering any fresh evidence. If you believed that the level of global worry about climate change was insufficient before that extrinsic cause for increased worry, you might decide that the level of collective worry after the movie better fits the magnitude of the threat. The movie does not need to offer fresh evidence to achieve this good effect. The movie compensates for climate change's deficit in charisma.

Suppose we apply Elke Weber's framework for allocating worry to artificial superintelligence. The impressive achievements of generative AI do seem to have caused an increase in discussions about this extinction threat. This would seem to suggest increased levels of collective worry about ASI. Is this increase justified?

What claim on our finite collective pool of worry should we grant to the scenario described by technologists and philosophers in which improvement in AI produces an artificial intelligence that rapidly surpasses us and resolves to send us extinct? How might we make an accurate assessment of how much of a claim on our finite pool of worry ASI deserves? We would need to evaluate how likely progress in AI is to produce one and how soon we should expect it. We could then balance it against other claims on our collective finite pool of worry.

We offer no such analysis here. Instead, our focus is on extrinsic factors – influences that have no direct bearing on how probable an unfriendly ASI is. We propose that these extrinsic factors may cause us to allocate a greater share of our finite pool of worry to ASI than it warrants.

What percentage of that increase results from intrinsic causes of change and what percentage is due to extrinsic causes? We conjecture that the accomplishments of OpenAI's ChatGPT do constitute an intrinsic cause for worrying a bit more. But some of that increase is due to extrinsic causes.

Consider the human-unfriendly AI in the 1991 movie *Terminator 2: Judgment Day*. In that movie, we learn that in 1997 the Skynet AI "becomes self-aware at 2:14 a.m. Eastern

time, August 29th" and then launches nukes at Russia. We cannot be certain that an AI product with those capacities is not scheduled for imminent release.

An AI with Skynet's capacities would be a radical departure from the Large Language Model AIs that OpenAI is mainly known for. OpenAI's LLMs are trained on large amounts of text from the internet. It is highly unlikely that the launch codes for America's nuclear arsenal exist in any form on the social news aggregation site Reddit, on Wikipedia, or on any webpage accessed by OpenAI's web crawlers. Nor is it likely that they can be inferred from that information.

We are, of course, free to imagine an ASI with such power that it could infer the identities of individuals with access to the launch codes for America's nuclear arsenal. In a logically possible thought experiment, once these individuals are identified, the conjectured ASI could use its superintelligence to divine the codes. In the age of surveillance capitalism, we are accustomed to the unnerving inferences that advertisers can make about our purchasing intentions based on our online behaviour. (Zuboff, 2019) We are, of course, some distance from a situation in which an AI could infer nuclear launch codes from the behaviour of those presumably rare individuals that know them with the ease with which Meta can conclude that posts on Facebook about politics suggest an interest in purchasing Harley Davidson motorcycles (Stephens-Davidowitz, 2017).

5. Viral Narratives about Extinction

The starting point for our discussion about extrinsic causes is the recent work of economist Robert Shiller (2019) on how narratives shape our economic choices. Shiller responds to a view in economics of people dispassionately calculating potential economic returns before they act. In Shiller's *Narrative Economics*, simplified and easily understood and transmitted economic narratives play a large role in motivating our economic choices. He describes how these narratives can go viral, amplifying their impact. One simple narrative about the welfare of the economy exaggerates the performance of a nation's stock market. This narrative leads us to focus too much on the stock market when deciding how well our nation is doing and not enough on indicators in other parts of the economy. Another powerful narrative points to cryptocurrency as the future of money. Shiller conjectures that this sustains investment in crypto through various scandals that should prompt caution, especially among naive investors.

We think that something similar occurs in respect of our fears about human extinction. Some narratives exercise a strong influence on the way we respond to AI. We see the presence of the Skynet narrative in much of the current discussion about ASI. In October 2023, we were treated to Elon Musk lecturing UK Prime Minister Rishi Sunak about the possibly imminent arrival of artificial superintelligence, and the latter wondering how political leaders could possibly regulate such a thing. These concerns were introduced by reference to Skynet in the *Terminator*. The movie franchise comes readily to mind whenever a journalist questions a tech visionary about the future. It tends to crowd out serious consideration of less charismatic consequences of the boom of interest in AI. These include the displacement of workers by AI and the environmental consequences of the power needed to train the latest AI models.

6. Comparing Two Charismatic Extinction Threats

When we consider possible causes of human extinction, we should understand artificial superintelligence as a charismatic extinction threat. Shiller would say that narratives about ASI go viral quite easily. Another way to make that point is that we make movies about ASI, and they do well at the box office. Few people who hear warnings about the potential of progress in AI to create Skynet respond, "What's Skynet?".

Consider the extinction threat featured in the post-apocalyptic drama television series *The Last of Us*. In that series, mass infection by a mutated Cordyceps fungus sparks a global pandemic that seems set to drive humanity to extinction. *The Last of Us* may be great TV, but that doesn't justify a significant claim of extinction-by-fungus on our global pool of worry. We should be aware of the ways in which our tendencies to find some extinction threats charismatic, and others not, can be influenced by extrinsic causes like sci-fi.

The Last of Us prompted much discussion about how likely or possible it would be for the Cordyceps fungus to infect humans. Suppose those wondering how much of their finite pool of worry they should allocate to this new extinction threat seek advice from *Scientific American*. They might find a discussion about the *Ophiocordyceps* genus, of which the fungus in *The Last of Us* is an example, that informs them that "no *Ophiocordyceps* species invades any fish, amphibians, or mammals" (Parshall 2023). The series motivates the threat by presenting unprecedented human infections as a consequence of climate change. A scenario in which a warming planet enables human infections violates no law of physics or logic. Perhaps a warming planet has increased the probability of such an event.

How much should we worry about extinction by Cordyceps? We conjecture that if we pose the question in terms of legitimate claims on our collective finite pool of worry, the answer is not much. If we were to worry a lot about extinction by Cordyceps, we may find ourselves also having to worry a great deal about imminent extinction by extraterrestrial invasion. The 2030 arrival of an extraterrestrial battle fleet cloaked from our primitive sensors violates no law of logic or physics that we know of. Well-acted dramas with high production values will continue to increase the charisma of extinction by fungal pandemic or extraterrestrial battle fleet. These extrinsic causes of increased worry offer little justification for increasing that worry.

What might count as an intrinsic cause for worrying more about an extinction threat? Suppose a well-researched scientific paper offered evidence that warming of the planet is making it easier for fungi of the *Ophiocordyceps* genus to overcome mammals' resistance to infection. The paper documents infections in mice. That would count as an intrinsic and therefore rationally defensible justification for increasing the claim of Cordyceps as a potential cause of human extinction on our collective finite pool of worry.

We should consider influences on the extinction threat from ASI in this light. The fact that a human-unfriendly ASI is possible suggests that it has a legitimate claim on our collective finite pool of worry. We have not addressed the intrinsic factors that would quantify the legitimate claim of ASI on our worry. Instead, we have pointed to extrinsic

factors that tend to change the amount we worry about extinction threats. We should acknowledge that ASI is a very charismatic cause of human extinction and that we have a propensity to overestimate its claim on humanity's pool of worry.

7. In Search of a More Rational Way to Allocate Our Concern about AI

Suppose charisma is a significant driver of our collective concerns about extinction. A downside of allocating too much worry to charismatic extinction threats is not allocating sufficient worry to other, less charismatic dangers. If we are overly worried about the threat from the charismatic Skynet and Cordyceps what uncharismatic threats to our species are we overlooking?

Climate change seems to be a paradigm of an uncharismatic extinction threat. For people in the rich world, it seems mostly to involve unexpectedly bad weather and unpleasant turbulence on business class flights to exotic holiday destinations. The greatest harms seem to be borne mainly by the poor. We didn't care much about them when the effects of climate change were not apparent. Now we find that their levels of suffering are still greater. As people in the rich world witness a decline in the high quality of their existences, what sacrifices are they prepared to make to prevent the poor from sliding into even greater levels of misery?

An anonymous referee suggested that Hollywood has, at times, helped elevate concern about climate change closer to the rational and moral optimum. *The Day After Tomorrow* (2004) dramatized climate change as a fast-paced disaster, depicting a tipping point that triggers an abrupt Ice Age. While such a scenario is not seriously considered by climate scientists, the film succeeded in boosting the charisma of climate crises more broadly.

We write these words as the world is emerging from the COVID-19 pandemic and wants to expeditiously move on from the tedium and anger of its many lockdowns. Steven Soderbergh's 2011 movie on a global pandemic *Contagion* performed well in the box office and received many awards. It served to focus attention on potential risks from a pandemic at that time. Remakes in the wake of the actual pandemic are likely to encounter a different market. Might the deficit in active worry about how to respond to outbreaks of contagious disease leave us tragically ill-prepared for the next pandemic? One lesson we should have learned about COVID-19 is that good responses require more than effective vaccines. It is a problem as we anticipate the responses that will be required by the next pandemic that pandemics have become anti-charismatic. We would rather contemplate which of our available weapons technologies might be effective against a T-800 Terminator cyborg than seriously consider under which circumstances it might be prudent to respond to a disease outbreak by going into lockdown.

How do we identify potential causes of extinction that we overlook, much in the way that we currently overlook the uncharismatic *Spalax arenarius*? If we want to optimize the potential for our collective pool of worry to give us timely warning of extinction threats, then we need methods that help us to identify a wide range of uncharismatic extinction threats that are just as capable of driving us to extinction as Skynet. We should mitigate

the propensity for Skynet and Cordyceps to take too great a share of our finite pool of worry.

We conclude with a suggestion about what we, as humanities scholars, can offer that does not falsely present us as experts on AI. Our discussion suggests that we are easily influenced by extrinsic factors in our assessment of risk. This means that we find it too easy to focus on charismatic extinction threats. Humanities scholars are well-positioned to bring to light low-charisma extinction threats that deserve more worry. This method should enable us to see past the glare of Skynet and Cordyceps.

Our point can be made by means of an insurance analogy. If you are purchasing insurance for a holiday, you can be over-insured against terrible mishaps. One way for your holiday to go horribly wrong would be if you were abducted by terrorists. But before you pay for insurance covering a team of mercenaries to rescue you, you should consider how much it is worth paying for such coverage. If your holiday is in a location with no history of terrorism, then perhaps you should forgo it and accept the small risk. A widely watched movie about holidaying innocents abducted by terrorists would be an extrinsic cause of increased fear of terrorism. We allow that it can be good for businesses selling coverage to encourage people to pay more for insurance than the actual risk warrants. As your potential insurer, we would look at the high price you would pay for terrorist coverage and immediately try to sell you coverage for extraterrestrial abduction—another misfortune with a non-zero probability that violates no law of logic or physics that we know of.

What we propose points to the imagination, a human attribute especially valued in the humanities. We suggest that we need *imagination insurance* to address potential causes of human extinction.

Imagination insurance offers a way to bring attention to extinction threats that overly focusing on Skynet and other exotic extinction threats causes us to overlook. We can think of this imaginative exercise as analogous to the stage an insurance provider must go through before it quantifies the risk and sets the premiums. The insurer of your holiday must first work out as many of the ways in which holidays can go very wrong. That list includes a wide variety of boring mishaps ranging from loss of luggage, through the impact of bad weather on travel, to sickness. It might also include some exotic causes of vacation mishap. An insurer that sells protection against abduction by terrorist or extra-terrestrial probably expects to profit from the charisma of these ways in which holidays can go wrong.

Humanity needs imagination insurance in respect of extinction threats. Rather than paying monetary premiums, we get imagination insurance simply by thinking expansively about our species' many possible futures. This should help us to identify causes of extinction, many of which are less charismatic than extinction by ASI. But they are causes of extinction, nonetheless.

What is the best way for us to anticipate and prepare for as many extinction threats as feasible? Earlier we mentioned the 8.1 billion minds that humanity potentially has available to worry about existential threats. Suppose we committed to engage in a

discovery phase of humanity's engagement with extinction threats. Before we get busy assessing the intrinsic factors that ought to tell us how much of our finite pool of worry we should allocate to each extinction threat, we need as complete a list of as many possible causes of extinction before we find ourselves staring at the proverbial asteroid about to make impact with Earth. Our 8.1 billion minds encompass many different ways of thinking about a tragic end for humanity. We need to maximize the imaginative reach of these billions of minds. Only then can we decide which threats to ignore and which to seriously prepare for.

One of us (Nicholas Agar) witnessed the power of imaginative diversity during a 2019 visit to Tongatapu, the main island of the Kingdom of Tonga. People were asked to think about their future and how to prepare for it. They engaged in a way that was entirely unfamiliar to Agar, trained in the ways of the Western academy. If we are engaged in the discovery phase of potential causes of human extinction, we should leverage the full reach of humanity's imaginative diversity. What ways for humanity to go extinct can the imaginations of Tonga access that are not so easily imagined elsewhere? Another way to express this question might be: What forms of apocalyptic sci-fi might be produced by the minds of Tonga that are less likely to be produced elsewhere?

Perhaps here is the true cost of the monoculture of stories that Hollywood has created. There is much discussion about Hollywood's contribution to America's soft power. Perhaps that is good for America's global influence. But there is a downside to the tendency for Hollywood's stories to become the world's stories. It tends to narrow our species' imaginative reach.

Chris Taylor's 2014 book *How Star Wars Conquered the Universe* charted the emergence and near-universality of the *Star Wars* stories (Taylor, 2014). Even people in remote communities find the shape of a *Star Wars* X-wing fighter familiar and know who Luke Skywalker's father is. Perhaps the near-universality of stories like *Star Wars* and the *Terminator* facilitates communication across barriers of language and culture. But there is a downside to this imaginative monoculture. When we are all obsessing about Skynet, how many other extinction threats does that focus cause us to overlook? If we are all dreaming of electric sheep, we risk a monoculture of the imagination in which we can dream only of Sith Lords, blade runners, and T-800 Terminator cyborgs.

These times of rapid technological change are generating a great deal of science fiction. We have offered no critique of this attempt to prepare for the future. We need the *Terminator* and *Star Wars*. It is good that fiction can take us beyond the limits of what today's technologies allow. Our complaint has focused on the propensity for charisma to grant some stories about the future too great a claim on our collective concern. If we challenge the charisma of the Terminator and the Death Star, what other threats might humanity's vast and varied imagination bring to light? How might they better prepare us for the real challenges we face?

Funding: This research received no external funding. But Murilo Vilaça thanks to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the support: APQ/PRÓ-

HUMANIDADES (421523/2022-0), APQ/Chamada Universal (421419/2023-7) and Bolsa de Produtividade em Pesquisa/PQ (315804/2023-8).

Acknowledgments: We are grateful to an anonymous referee for very helpful suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

Agar, N. (2016). Don't Worry about Superintelligence. *Journal of Ethics and Emerging Technologies* 26(1): 73-82.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Courchanp F. et al. (2018). The Paradoxical Extinction of the Most Charismatic Animals. *PLOS Biology* 16(4): e2003997.

Kean, S. (2025). The Comet Panic of 1910, Revisited. *Science History Institute*. Available at: https://www.sciencehistory.org/stories/magazine/the-comet-panic-of-1910-revisited/

Keil, F. (2006). Doubt, deference, and deliberation: Understanding and using division of cognitive labor. In *Oxford Studies in Epistemology*. Tamar Szabo Gendler and John Hawthorne (eds.). Oxford: Oxford University Press, 2006, pp. 143-166.

Lynas, M. et al. (2021). Letter to Environmental Research "Greater than 99% consensus on human caused climate change in the peer-reviewed scientific literature" *Environ. Res. Lett*. 16 114005DOI 10.1088/1748-9326/ac2966

Mukherjee, S. (2010). *The Emperor of All Maladies: A Biography of Cancer*. New Yor: Scribner.

Parshall, A. (2023). Could the Zombie Fungus in TV's The Last of Us Really Infect People?. *Scientific American*. Available at: https://www.scientificamerican.com/article/could-the-zombie-fungus-in-tvs-the-last-of-us-really-infect-people/

Shiller, R. (2019). *Narrative Economics: How Stories go Viral and Drive Major Economic Events*. Princeton University Press.

Stephens-Davidowitz, S. (2017). *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*. Dey Street Books.

Taylor, C. (2014). *How Star Wars Conquered the Universe*. New York: Basic Books.

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf.

Vilaça, M. & Lavazza, A. (2022). Not Too Risky: How to Take a Reasonable Stance on Human Enhancement. *Filosofia Unisinos* 23(3): e23305.

Weber, E. (2010). What Shapes Perceptions of Climate Change? *Climate Change* (1): 332-342.

Weber, M. (1947). *The Theory of Social and Economic Organization*. Oxford: Oxford University Press.

Yampolskiy, R. (2015). *Artificial Superintelligence: A Futuristic Approach*. New York: Chapman and Hall/CRC Press.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Profile Books.