WORKING PAPER / 2023.03

Effective Altruism
Doing transhumanism better

Mollie **Gleiberman**

University of Antwerp
IOB | Institute of Development Policy

The IOB Working Paper Series seeks to stimulate the timely exchange of ideas about development issues, by offering a forum to get findings out quickly, even in a less than fully polished form. The IOB Working Papers are vetted by the chair of the IOB Research Commission. The findings and views expressed in the IOB Working Papers are those of the authors. They do not necessarily represent the views of IOB.

Institute of Development Policy

Postal address:
Prinsstraat 13
B-2000 Antwerpen
Belgium

Visiting address:
Lange Sint-Annastraat 7
B-2000 Antwerpen
Belgium

Tel: +32 (0)3 265 57 70
Fax: +32 (0)3 265 57 71
e-mail: iob@uantwerp.be
http://www.uantwerp.be/iob

WORKING PAPER / 2023.03

ISSN 2294-8643

Effective Altruism
Doing transhumanism better

April 2023

Mollie **Gleiberman***

* PhD student, University of Antwerp. Author contact: dmgleiberman@gmail.com.

This working paper is a chapter excerpted from a larger, forthcoming study about EA; comments are welcome.

University of Antwerp
IOB | Institute of Development Policy

Effective Altruism: Doing Transhumanism Better

Mollie Gleiberman[1]

Abstract: Effective Altruism (EA) is a Trojan horse for transhumanism, through which EA movement leaders and funders aim to naturalize transhumanism as the logical extension of the existing global aid and development sector. This paper traces transhumanism's mainstreaming, first via its rebranding as a humanitarian effort to save lives, protect vulnerable populations, and ensure global flourishing (what I term 'transhumanitarianism'), and later by embedding transhumanitarianism in EA (now under the rubric of 'longtermism'). A key component of this strategy was inverting transhumanism's techno-optimism to instead focus on safety and preventing existential risks ('x-risk') *from* emerging technologies like AI and biotechnology, while simultaneously advocating *for* the creation of these same technologies. The paper focuses on some components of this strategy: the use of inoculation, speculative ethics, anticipatory governance, and the mobilization of apocalyptic discourse as means for producing material outcomes in the form of policy and research agendas.

[1] PhD student, University of Antwerp. Author contact: dmgleiberman@gmail.com. This working paper is a draft chapter excerpted from a larger, forthcoming study about EA; comments are welcome. I am grateful to Apolline Taillandier for helpful comments on an earlier version. All remaining errors are my own.

1. Introduction

Despite being commonly defined as an evidence-based approach to philanthropic giving that focuses on addressing global poverty, Effective Altruism (EA) is, in practice, an ideological and interest-driven project whose main aim is steering research and policymaking related to emerging technologies, particularly artificial intelligence (AI) and biotechnology. This agenda, which the movement now calls 'longtermism', reflects the *ideological aims* of EA's founders (members of an initially online subculture that coalesced in the mid-2000s around transhumanist thinkers Eliezer Yudkowsky, Nick Bostrom, David Pearce, Robin Hanson, and Aubrey de Grey) and the *financial interests* of the movement's major funders and supporters (Silicon Valley tech billionaires invested in AI/machine learning, biotechnology, cryptocurrency, and prediction markets). 'Longtermism' is the EA idea that since the future holds so many more people than the present, efforts to maximize wellbeing and reduce suffering (to 'do the most good') ought to prioritize the welfare of the entire aggregate future of humanity, and ensuring the well-being of the long-term future should be *a* —if not *the*— key moral priority of our time (EA Forum, 2021; MacAskill, 2019; Todd, 2018, 2019). Crucially, in terms of practical components, EA's 'longtermist' agenda is not merely *similar* to transhumanism, but precisely what Bostrom, Yudkowsky and their fellow transhumanists (later folded into a movement known as the 'rationalists') began advocating in the late 1990s and early 2000s. The outer justification has changed, jettisoning the unbridled techno-optimism that characterized the earlier Extropian transhumanism in favor of sober calls for safety and global well-being, but the core agenda—from the futuristic goals of space colonization, superintelligent artificial intelligence (also known as artificial general intelligence, AGI), genetic and cognitive enhancement, paradise engineering, and digital minds to the more down-to-earth goals of building 'civilizational refuges', popularizing prediction and forecasting markets, promoting 'rationality', cryptocurrencies and charter cities—remains strikingly the same[2].

While the specific content of this agenda is interesting and worthy of extended analysis in its own right, my aim here is more modest: I seek to trace the arc of transhumanism's mainstreaming via its positioning as the logical extension of the global aid and development sector — first as a humanitarian effort to save lives, protect vulnerable populations, and ensure global flourishing and well-being (what I term 'transhumanitarianism'), and later through EA (under the rubric of 'longtermism'). A key component of this strategy was inverting transhumanism's techno-optimism to instead focus on safety and preventing existential risks ('x-risk') to the future of humanity from emerging technologies, while advocating for the creation of these same technologies. This process of *inoculation*—pre-emptively admitting the flaws of whatever ideology, project, or worldview one is promoting in order to protect and strengthen it from attack (Mosco, 2005)—combined with prognostications of catastrophe has enabled the advancement of the transhumanist sociotechnical imaginary to ascend global policy and research agendas. The paper focuses on the tactical components of this mainstreaming: the use of inoculation, speculative ethics, anticipatory governance, and the mobilization of apocalyptic discourse as means for producing material outcomes in the form of policy and research agendas.

2. Transhumanism

As explained by Nick Bostrom, one of the leading proponents of contemporary transhumanist thought, transhumanism is 'an outgrowth of secular humanism and the Enlightenment' based on the idea that 'current human nature is improvable through the use of applied science and other rational methods' including currently existing technologies and potential emerging/future technologies, such genetic engineering, artificial intelligence, nanotechnology and fully-immersive virtual reality (Bostrom, 2011b, p. 55). Broadly, transhumanists posit that 'ethical problems frequently have technical solutions' (Pearce, 2010). These technical solutions typically involve imagined technologies that have *not yet been*

[2] For a quick comparison, see (Bell & O'Connor, 1988; Extropy Institute, 2003), which list the Extropian transhumanist goals, and compare this to, e.g., the types of projects promoted by the FTX Future Fund, led by two co-founders of the EA movement (FTX Future Fund, 2022a, 2022b).

invented but which transhumanists *hope* will be developed; hence much of the transhumanist literature aims to drive interest and funding toward developing these technologies (Hall, 2017; Hauskeller, 2016; Tirosh-Samuelson, 2011). Key among these are the development of artificial intelligence (AI) and the Singularity[3]; space colonization; human and non-human cognitive, moral, and physical enhancement; and the elimination death via cryonics, mind-uploading, and life-extension biotechnologies (Bostrom, 2005b; Sandberg, 2015). Effectively, transhumanists desire to fundamentally alter the trajectory of life on Earth (and beyond) and direct the evolution of human and non-human species on our planet, fulfilling their dream of achieving 'technological maturity'[4] and colonizing space (Bostrom, 2003a, 2008a; Matheny, 2006). To ensure the fulfillment of this destiny, they work to ensure the creation of radical biotechnologies and superintelligent AI (also known as artificial general intelligence, AGI), the latter of which will first solve mankind's most pressing problems—poverty, climate change, illness, death—for us, and then merge with us in some glorious post-human form and spread throughout the galaxy, perhaps even as pure intelligent energy, aka, the 'Omega Point' (Barrow & Tipler, 1986; Moravec, 1988).

Despite rejecting theism of any kind as *irrational*, the transhumanists' own project is saturated with religious symbolism that echoes apocalyptic and millenarian thinking, none more-so than the overtly eschatological connotations of the Singularity (Geraci, 2010; Hauskeller, 2016; Pinto, 2019; Ranisch & Lorenz Sorgner, 2014; Tirosh-Samuelson, 2014; Tirosh-Samuelson & Hurlbut, 2016), which, transhumanists believe, will either usher in a future of extraordinary flourishing and the elimination of all suffering, or a dystopia that leads to human extinction (Bostrom, 2001). Wedded to a discourse of technological-progress-as-salvation (Burdett, 2015; Verdoux, 2009), transhumanists argue that we have a *moral imperative* to pursue human (and non-human) enhancement using technology (Bostrom, 2005a; de Grey, 2006b, 2007; Harris, 2009; Nuland, 2005; Pearce, 2007; Savulescu & Sandberg, 2008). In the extreme utilitarian logic of the transhumanists, to delay creating these technologies (or to fail to create them *at all)* is to consign millions of potential future living beings to substandard lives, or worse, to non-existence: a catastrophic, '*astronomical waste'* of potential value (Bostrom, 2003a). Portraying the wondrous post-human future that awaits humanity (see, e.g., Bostrom, 2006; Bostrom, 2008b) as dependent upon an epistemic and cognitive victory over contemporary beliefs and norms that reject the transhumanist worldview (Bostrom & Ord, 2006; Verdoux, 2009) the transhumanists' most urgent task is to *proselytize*: to convince others of the fundamental rightness of their worldview and ensure 'emerging technologies' are prioritized on global research agendas. Doing so requires shifting the Overton window[5] of socially and politically acceptable norms, values, and beliefs about fundamental metaphysical questions, such as what constitutes a well-lived life, how far technology should intervene in human evolution, humanity's place in the Universe, and the very meaning of life and death.

3. Transhumanitarianism

The idea of using technology to transcend human limitations has a deep history spanning centuries, but contemporary transhumanism is typically traced to the subculture that coalesced around Max More's Extropianism in the 1980s (Bostrom, 1999). The Extropian transhumanists of the 1980s and 1990s were decidedly libertarian, often explicitly hostile to the idea of government, and styled themselves as a radical counter-culture advocating for the accelerated development of new technologies to transcend the human condition (Bell & O'Connor, 1988) — a politics forged from San Francisco bohemianism

[3] The Singularity refers to the moment at which artificial intelligence becomes equal to or surpasses human intelligence; in other words, the creation of greater-than-human intelligence. In recent years, the phrases 'intelligence explosion' and 'artificial general intelligence' (AGI) have replaced references to the Singularity to create distance from the fringe rhetoric of the early singularitarians.

[4] Technological maturity is defined as 'the attainment of capabilities affording a level of economic productivity and control over nature close to the maximum that could feasibly be achieved' (Bostrom, 2013, p. 20).

[5] The Overton window refers to the window of political possibility, based on the idea that options available to policymakers are limited to those which already have a level of acceptance in wider society. This means that social institutions that shape norms, values, and ideas play a key role in establishing the range of policy options. For an idea to become a viable policy option, it must first be accepted as thinkable, reasonable. (Mackinac Center, n.d.)

and Silicon Valley technolibertarianism, famously christened 'the California Ideology' (Barbrook & Cameron, 1996).

By the late 1990s, however, transhumanism's affiliation with radical libertarianism and unbridled techno-optimism was becoming a liability for those who disagreed with Extropian politics (Bostrom, 1999), and wanted transhumanism to be taken seriously as a *philosophy* (Miller & James, 2006) and as a *progressive project* advocating the safe development of technology for the benefit of humanity. Transhumanists like Nick Bostrom, James Hughes, David Pearce and others distanced themselves from the rhetoric of the Extropians by creating new organizations, such as the World Transhumanist Association (WTA; now Humanity+), the Institute for Ethics and Emerging Technologies (IEET), and the Future of Humanity Institute (FHI), which is housed at Oxford University. Their goal was to turn transhumanism—till then considered the realm of internet cranks and science fiction junkies—into a serious topic for academics and policymakers (Bostrom, in Humphrey, 2004).

In an important strategy document, Hughes (2002) suggested that transhumanists were likelier to attract the necessary public and political support to achieve their goals if they concentrated on ensuring the *safety* of emerging technologies and affirmed a commitment to democratic values of equity and fairness. To successfully challenge critics ('bio-luddites') and gain mainstream credibility, Hughes argued, the transhumanists had to reject their 'elitist anarcho-capitalist roots' and instead 'embrace the need for government action to ensure that transhuman technologies are safe, effective, and equitably distributed' (Hughes, 2002). Put simply, transhumanism would be rebranded and presented as a sensible philosophical movement concerned with promoting the ethical use of technology for the benefit of all. This entailed situating the pursuit of radical new technologies within existing frameworks of *sound science* and the *pursuit of the greater good*: a deeply humanitarian effort, concerned with addressing inequality, saving lives, ending suffering, and promoting a flourishing future for all. I will refer to this effort to cast transhuman objectives as a humanitarian, life-saving mission as *transhumanitarianism*.

Transhumanitarianism developed along two main axes: *life-saving transhumanitarianism* (dedicated to making the case for transhumanist technologies as life-saving/life-enhancing interventions in and of themselves, and stressing benefits and opportunities), and *x-risk transhumanitarianism* (focused on safeguarding the welfare of future generations, particularly by preventing existential risks, or 'x-risks', to the future of humanity from emerging technologies like AI, and stressing threats and catastrophic ruin). In both cases, utilitarian ethics towards future populations and 'optimal philanthropy'—ensuring the greatest number of lives saved or improved per philanthropic dollar donated—are employed to depict transhumanism as a charitable cause meriting generous philanthropic support.

Life-saving transhumanitarianism focuses on the opportunities that transformative technologies like superintelligent AI, cryonics, and enhancement would bring if invented, positing them as the ultimate solutions to wicked problems: '[d]isease, poverty, environmental destruction, unnecessary suffering of all kinds; these are things that superintelligence equipped with advanced nanotechnology would be capable of eliminating' (Bostrom, 2003b). It would be far more cost-effective to focus on creating smarter-than-human AI, 'as a means of directly solving such contemporary problems as cancer, AIDS, world hunger, poverty, et cetera' (Singularity Institute for Artificial Intelligence, 2002b). Indeed, donors seeking to do the most good with their charitable dollar would be hard-pressed to find a more effective cause than Yudkowsky's efforts to create superintelligent AI:

> Is there anywhere else where a small donation would do more good? The Singularity is a tremendously effective means of addressing human problems. […] The Singularity is the most effective means we know for investing a given amount of money so that it brings the largest possible amount of real good to the greatest number of people. (Singularity Institute for Artificial Intelligence, 2002a)

Or, for those unsure of the Singularity, Aubrey de Grey's project to end death by curing aging also promised to save more lives than traditional charitable causes:

> [S]aving lives is the most valuable thing anyone can spend their time doing […] since over 100,000 people die every single day of causes that young people essentially never die of, you'll save more lives by helping to cure aging than in any other way. (de Grey, 2006a)

Addressing the question of whether it is 'more urgent to feed those who are starving today', de Grey argued that prioritizing saving lives in the present over funding anti-aging research denies the intrinsic value of those living in the future: 'even if there were a choice between feeding the starving and curing aging, the arithmetic of healthy years added to people's lives by the two policies […] argues that we should put most of our effort into curing aging' (de Grey, 2006a).

Life-saving transhumanitarianism endeavored to create distance from the Extropians' individualistic rhetoric by emphasizing collective well-being. However, it still openly championed transhumanist technologies, and thus remained vulnerable to the charge of rampant techno-optimism. A more effective strategy was to take the opposite tack: focusing on preventing existential risks ('x-risk') from emerging technologies and emphasizing safety. In contrast to the optimism of life-saving transhumanitarianism, this *x-risk transhumanitarianism* stresses the urgent need for safety and caution by highlighting the existential risks transformative technologies pose — including (somewhat paradoxically) the risk that such technologies might *not* be invented, preventing humanity from reaching 'technological maturity'. This emphasis on safety was rapidly adopted by transhumanists, e.g., Yudkowsky's quest to bring about a positive singularity ('Friendly AI') was replaced by an effort to ensure the *safe* development of AI ('AI-safety'). The lodestar in this effort was Nick Bostrom's "Astrononomical Waste" argument, to which I turn next.

4. Astronomical Waste: Speculative Ethics Meets Anticipatory Governance

Bostrom's "Astronomical waste" paper (Bostrom, 2003a) argues that since the future could contain potentially vast numbers of people, the majority of total aggregate 'value' in the universe (from a total utilitarian perspective) lies in the future[6]. When space colonization is factored in, thus allowing for the continued survival of humanity after Earth becomes uninhabitable due to the death of our sun, that potential future value is *even greater*. Any delay in space colonization—even by just one second—means the loss of an enormous number of potential lives, something that utilitarians who care about maximizing total value should be very concerned about:

> Advancing technology (or its enabling factors, such as economic productivity) even by such a tiny amount that it leads to colonization of the local supercluster just *one second* earlier than would otherwise have happened amounts to bringing about more than 10^{29} human lives […] that would not otherwise have existed. Few other philanthropic causes could hope to match that level of utilitarian payoff. (Bostrom, 2003a, p. 4)

Bostrom notes that this would seem to indicate that philanthropic utilitarians ought to focus solely on accelerating technological development to ensure space colonization occurs as soon as possible, since the 'payoff from even a very slight success in this endeavor is so enormous that it dwarfs that of almost any other activity' (ibid., p. 5). However, he continues,

> the true lesson is a different one. If what we are concerned with is (something like) maximizing the expected number of worthwhile lives that we will create, then in addition to the opportunity cost of delayed colonization, we have to take into account the risk of failure to colonize at all. We might fall victim to an *existential risk*, one where an adverse outcome would either annihilate Earth-originating

[6] In debates about intergenerational justice and obligations to future generations (see, e.g. Parfit 1984; Pasek, 1992), philosophers regard this view as an example of *temporal impartiality*: if we agree that all people are of equal moral worth, this view argues, then we are obliged to extend moral consideration equally to all future people (Walker, 2007). JJC Smart, whose discussion of utilitarianism and the far future concisely outlines many of the planks of contemporary transhumanism, argued that to deny the equal value of future generations (such as by using discount rates, as in economic calculations about future risks) was to be 'temporally parochial' (Smart, 1973, p. 63). Some philosophers have argued that not only should future generations be considered equally, but they should perhaps be entitled to preferential treatment: first, since they will greatly outnumber us (in aggregate), and second, because they can be classified as far more vulnerable than any current population given that they have no voice in current debates and decisions (Petrucci, 1998, p. 50).

> intelligent life or permanently and drastically curtail its potential. […] For standard utilitarians, priority number one, two, three and four should consequently be to reduce existential risk. The utilitarian imperative 'Maximize expected aggregate utility!' can be simplified to the maxim 'Minimize existential risk!'. (Bostrom, 2003a, pp. 5-6)

Bostrom thus sets out two seemingly incongruent paths for securing the future: one that focuses on the importance of *advancing technological progress* (to ensure that we create all the technologies, such as AGI, that will allegedly enable us to colonize space sooner rather than later), and one that focuses on *preventing existential risks* ('x-risks') which would prevent space colonization from ever being achieved (which would, for transhumanists, mean the loss of all value in the universe). Bostrom clarifies that while some x-risks are natural in origin (asteroid strikes, supervolcanoes, natural pandemics), he believes we should be far more worried about x-risks and 'global catastrophic risks' that are *anthropogenic,* and which stem from our technological capacity to destroy ourselves (Bostrom, 2002, 2011a, 2013, 2014; Bostrom & Ćirković, 2008; Bostrom & High, 2016). Familiar examples of anthropogenic risks are nuclear warfare and climate change, but, Bostrom argues, the *most pressing* anthropogenic threats are from emerging technologies such as superintelligent AI, bioengineering/synthetic biology, and nanotechnology (Bostrom, 2003c, 2014; Bostrom et al., 2016).

Of course, these are precisely the same family of emerging (or 'transformative') technologies that transhumanists believe are necessary for 'civilization' to *survive and flourish*, as opposed to *stagnate and decline*. This apocalyptic narrative serves as the justification for particular material interventions (Koch, 2021), namely, the *safe* development of emerging technologies. Bostrom's framing deftly inoculates (Mosco, 2005) transhumanism's techno-myth by encasing his *advocacy for* emerging technologies within expressions of *concern about* their safe development, arguing that unsafe development would lead to catastrophe or even extinction. Invoking what Whyte (2021) refers to as *crisis epistemology*—the exercise of power that extends colonizing logic, justified through claims of offering the solution to an immediate and unprecedented situation—the transhumanists mobilize the specter of future ruin to validate the imposition of their sociotechnical will upon the world. Catastrophe is invoked to open space for crisis management, empowering those who present themselves as offering the solution with the right to act (Swyngedouw, 2013): developing *safe* AI becomes a *global priority*. Despite being framed as protection from extreme vulnerability, Schuster and Woods (2021) argue, 'Bostrom is more interested in smoothing the way toward a future superintelligent existence at cosmological scales than examining current risks and rewards of being a precarious human subject' (p. 91).

Transhumanists like Bostrom intervene in existing debates about, e.g., the misuse of AI and biotechnology, environmental destruction, and nuclear weapons, and appropriate them for their own purpose. Concerns about the dangers posed by emerging technologies are shared by many people—your author included—but Bostrom and his followers are ultimately motivated by very different reasoning than, say, AI ethicists, climate change activists, or policymakers working on the non-proliferation of nuclear/biological/chemical weapons (which often leads transhumanists to be dismissive of more immediate concerns, particularly those raised by AI ethicists). From transhumanism's perspective, if badly designed superintelligent AI or a bioengineered pandemic causes human extinction, then humanity will never reach 'technological maturity'[7]: Earth-originating intelligence will never colonize the galaxy, and trillions of *potential* lives ('value', 'utility') will be lost. Bostrom's concern is framed to leapfrog over present debates about whether technologies like superintelligent AI constitute a worthwhile, meaningful, or even technologically-feasible goal; his

[7] Technological maturity is defined as 'the attainment of capabilities affording a level of economic productivity and control over nature close to the maximum that could feasibly be achieved' (Bostrom, 2013, p. 20). The notion of technological maturity is deeply deterministic, assuming that there is some finite list of all the technologies that are possible under the laws of nature, which humanity is currently making its way through and ticking off, like a shopping list that can finally reach completion.

framing naturalizes the development of AGI and radical biotechnologies as inevitable, hence requiring urgent attention and funding to ensure they are developed *safely* rather than *recklessly*.

The notion of *inevitability* does a considerable amount of heavy-lifting here, pre-empting critics who might suggest simply avoiding creating these potentially-lethal technologies, through bans, restrictions, and regulations. *Not* creating AGI is simply not an option, this view says: barring an extinction event that eliminates us before we invent AGI, AGI is coming — whether we want it or not — hence we must take initiative *now* to steer it in a beneficial direction:

> Failing some cataclysmic event that destroys us prematurely, it [superintelligent AI] will happen. So the question is, how can one try and ensure it happens under as favourable conditions as possible? (Bostrom, in Rees, 2018)

Confusingly, hidden in plain sight is the acknowledgment that the development of particular technologies may be compromised *if* obstacles arise that prevent their invention (tacitly undermining the presumption of inevitability that animates the urgency of the discourse), in the form of the extinction of humanity or other catastrophic event that halts technological progress. Or, society might simply collectively choose a different path and decide not to pursue the invention of these technologies (which is, paradoxically, an equally catastrophic outcome by transhumanism's lights). Some observers have mistakenly interpreted Bostrom's concerns about the existential risk from superintelligent AI (particularly as elaborated in his 2014 popular non-fiction book, *Superintelligence*) as a sign that Bostrom now advocates *against* the development of emerging technologies, but this is inaccurate; he believes 'it would be a huge tragedy if machine intelligence were never developed' since this would represent the failure of 'Earth-originating intelligent civilization' to achieve its potential (Bostrom, in Achenbach, 2015). Indeed, Bostrom and his supporters argue that never reaching 'technological maturity' via the creation of those technologies transhumanists deem essential to 'desirable future development' is *itself* an x-risk (Andersen & Bostrom, 2012).

Bostrom's signature discursive move consists of what Alfred Nordmann (2007) calls *speculative ethics*: 'casting remote possibilities or philosophical thought-experiments as foresight about likely technical developments' (p. 31). Nordmann argues that authors who engage in speculative ethics employ an *if-then* tactic that treats hypothetical, imagined futures as imminent: sliding from an 'if' (a speculative scenario) to a 'then' (the need for urgent policy-relevant decisionmaking), which lends credibility to the speculative scenario and, by elevating it to current policy agendas (if we accept that X is plausible/likely/imminent, then we must start doing Y to prepare for it) displaces other concerns that may matter more in the actually-existing context. The finite attention of policymakers, and the finite resources of governments and other funders, gets redirected toward entirely speculative scenarios, while more immediate but less dramatized concerns fall off the agenda (Nordmann & Rip, 2009).

Bostrom makes liberal use of the if-then tactic throughout his writing, glossing the speculative nature of his scenarios to argue that by dint of their being imaginable as theoretically plausible—breezily skipping over questions about whether they are *actually* possible—that they warrant treatment as pressing public policy concerns (Jones, 2009). Bostrom's argumentative technique entails 'a curious reversal of the burden of proof to promote the displacement of the present by a hypothetical future' (Nordmann, 2007, p. 39); this rhetorical tactic creates space for *anticipatory governance* (Anderson, 2007). Anticipatory governance relies on the logic of preemption, which 'brings the future into the present [and] makes the future consequences of an eventuality that may or may not occur indifferent to its actual occurrence' (Massumi, in Anderson, 2007, p. 159). This can be found across Bostrom's work, wherein he argues that since it is impossible to *prove* with certainty that a particular imagined future scenario will *not* come to pass, then we must assume that such imagined future scenarios are *plausible* and thereby worthy of attention and funding (Jones, 2009). In other words, if we cannot offer an irrefutable argument that something will *not* happen, we are (in this view) tacitly admitting that there is a reasonable chance it *will* happen. For instance, since we have no definitive proof that superintelligent

AI *will not* be invented within the next fifty years, Bostrom argues that we ought to begin planning as if it *will*:

> [T]here is currently no warrant for dismissing the possibility that machines with greater-than-human intelligence will be built within fifty years. On the contrary, we should recognise this as a possibility that merits serious attention. (Bostrom, 2003c)

5. Effective Altruism: From Astronomical Waste to Longtermism (With a Detour Through Global Poverty)

The most extraordinary vector of transhumanitarianism has been the creation of the Effective Altruism (EA) movement, which famously advocates for 'doing good better' (the title of the movement's most popular introductory text, published by EA movement co-founder Will MacAskill in 2015). EA advocates a utilitarian-inspired, cost-benefit approach to philanthropy and aid evaluation, focused on saving or improving the most lives per dollar. EA's popular, 'public-facing' content stresses that donors who care about maximizing their philanthropic impact should donate to causes serving the 'distant poor': people in very poor countries who are dying of easily treatable or preventable diseases like diarrhea and malaria, where a dollar goes further (e.g. saving lives in sub-Saharan Africa through low-cost, high-impact aid interventions such as anti-malaria bednets and deworming pills, the top recommended causes by EA's flagship charity evaluator, GiveWell). However, those who become more deeply involved in EA encounter a different message: if one *truly* cares about saving or improving the most lives possible, one ought to prioritize the entire future of humanity by ensuring the positive development of safe AI and the prevention of x-risks. Sound familiar? It turns out that EA's key intellectual architects were all directly or peripherally involved in transhumanism, and the global poverty angle was merely a stepping stone to rationalize the progression from a non-controversial goal (saving lives in poor countries) to transhumanism's far more radical aims (Gleiberman, 2023). EA is a Trojan horse for transhumanism, through which movement leaders and funders have attempted to naturalize the transhumanist agenda as the logical successor to the existing global aid and development sector.

Explicitly building from Nick Bostrom's work and his 'astronomical waste' paper, Effective Altruists (EAs)[8] argue that there are grave anthropogenic *existential risks* ('x-risk') and *global catastrophic risks* to the future of humanity in which AI that is unaligned with human values, the misuse of biotechnology, or another unprecedented (technological) disaster could lead to humanity's extinction (Centre for Effective Altruism, 2020; Todd & 80000 Hours Team, 2019). From a total utilitarian perspective, extinction means not only the loss of all current people, but the loss of all future people — a population that, in aggregate, dwarfs that which is living today (Beckstead, 2013; Todd, 2013). Framed as *protecting the welfare of future generations* and *safeguarding the long-term future* (language usually associated with climate change and environmental activism), EA discourse skillfully manipulates the concept of the 'distant poor' such that the *temporally distant* become those who *most* need our help:

> Many people believe that we should care about the welfare of others, even if they are separated from us by distance, country, or culture. The argument for the long term future extends this concern to those who are separated from us through time. Most people who will ever exist, exist in the future. (Centre for Effective Altruism, 2020)

Since EA's welfarist, utilitarian logic stresses impartiality (all persons have equal more worth), maximization (we should seek to help the greatest number possible), and aggregation (summing up all 'value'), safeguarding the welfare of future generations has become *a*, if not *the*, top EA priority (Todd, 2017). Future generations, the EAs argue, are not only the *largest* population, but the *most disenfranchised*, hence 'if our aim is to do the most good, we should focus primarily on the effects of our present choices on the very distant future (thousands, millions, or billions of years from the present)'

[8] Obviously, the EA movement contains many individuals, who hold diverse views and do not speak with a single voice; when I refer to 'EAs' here I am speaking of the predominant view within the EA social movement and community.

(Macaskill & Tarsney, 2019; parentheses in original). The movement has adopted the umbrella term 'longtermism' to describe this perspective:

> Longtermism is the idea that because such huge numbers of individuals might live in the long-run future, and because we think everyone's interests matter equally, approaches to improving the world should be evaluated mainly in terms of their potential for long-term impact – over thousands, millions, or even billions of years. (80000 Hours, 2021)

An important corollary to *preventing* x-risks and making sure the future does not go badly is the idea that we can take steps today to ensure a *flourishing* future for millions of years, by enacting 'technological or civilizational trajectory changes' (Greaves et al., 2019, p. 12) to steer humanity through this *especially* dangerous period (Centre for Effective Altruism, 2018; Forethought Foundation, 2018; Halstead, 2019; Harris, 2019), into a safe, stable position as a 'technologically mature', spacefaring civilization that has colonized the galaxy (Beckstead, 2013; Dickens, 2020; EA Forum, n.d.; Harris, 2019; Karnofsky, 2021a; MacAskill & Islam, 2020; Todd, 2020). These trajectory changes include positively shaping the development of AI (through 'AI-safety' and 'AI alignment' research) and biotechnology (through funding biosecurity and people who work on 'safe' biotechnology), while pushing government institutions and decisionmaking bodies to adopt EA's worldview (John, 2019). Exhibiting the historical amnesia common to techno-myths (Mosco, 2005), 'longtermist' ideology posits that we are living at a pivotal point in history —a 'hinge of history' (EA forum, 2022; Fisher, 2020), a 'precipice' (Ord, 2020), or 'the most important century' (Karnofsky, 2021b)— wherein actions we take today will determine whether humanity has a bright future colonizing space and filling the universe with value, or is damned to stagnation, decline, and extinction.

'Longtermism' is lauded by EAs as an important new worldview, intellectual project, academic field, and research paradigm; however, 'longtermism' is nothing more than a skillful rebranding of the transhumanitarianism discussed in the previous section — it is transhumanism, divested of its controversial origins and presented as the next frontier of global development. 'Longtermism' superficially tempers the techno-optimism of those earlier, radical futurists, focusing instead on the role of technological progress in welfare gains during the past few centuries (for instance, movement discourse now highlights the smallpox vaccine and the Green Revolution), while stoking fears along two fronts: one positing that emerging technologies such as AI and biotechnology *themselves* constitute a potential existential threat, and one positing that *failure to develop* these same emerging technologies would lead to stagnation, leading inexorably to extinction and the loss of all value in the universe.

The emphasis on progress vs. stagnation and the role of technology in ensuring 'civilizational survival' (and averting 'civilizational collapse') draw upon the celebratory discourse of *progress* advanced by thinkers like Tyler Cowen (Cowen, 2010, 2013), Patrick Collison (Collison & Cowen, 2019; Zuckerberg et al., 2019), Steven Pinker (Pinker, 2015), and Peter Thiel (Harrington, 2022; Ngo, 2020; Thiel, 2013; Thiel & Cowen, 2015; Thiel & Masters, 2014; Weinstein & Thiel, 2019). This discourse of progress reframes contemporary fears of civilizational collapse and dystopian futures *not* as resulting from the excesses of capitalism and the pursuit of exponential growth, but from their opposite. In the progress-centric view, calls for degrowth and skepticism toward the triumphalist discourse of (Western) technological progress could lead to permanent stagnation. In this view, learning to live within our planetary means constitutes a *threat* to the future of humanity: if humanity never escapes the planet, when the Earth becomes uninhabitable and our sun dies out, all Earth-originating sentience and intelligence will be permanently lost, depriving the Universe of a *valuer*, and hence, value. Thus, even as EA's turn toward 'longtermism' appears to be primarily focused on averting risks from emerging technologies, in practice, this is all undertaken in the service of a massive advocacy campaign for creating those very same technologies (even if many EAs fail to fully acknowledge this).

Following Bostrom, Cowen, Collison, and Thiel, EAs believe that halting or banning efforts to develop superintelligent AI and radical biotechnologies is wrong-headed: not only are transformative

technologies understood to be inevitable[9], but *not* developing them would constitute an existential catastrophe itself, since it would mean failure to reach our *potential*:

> [H]umanity never building AGI, never realizing our potential, and failing to make use of the cosmic endowment would be a tragedy comparable (on an astronomical scale) to AGI wiping us out. (Soares, 2018; f.8)

> Failing to reach technological maturity is also classed as threatening the future of humanity, even though it may not sound like a particularly awful scenario, because of the huge loss of potential. (Whittlestone, 2017; f.2, annotated quote from Bostrom in original)

> Perhaps the most concerning risk to civilization is that we continue to exist for millennia and nothing particularly bad happens, but that we never come close to achieving our potential—that is, we end up in a "disappointing future." A disappointing future might occur if, for example: we never leave the solar system; wild animal suffering continues; or we never saturate the universe with maximally flourishing beings. In comparison to civilization's potential, a disappointing future would be nearly as bad as an existential catastrophe (and possibly worse). (Dickens, 2020; parentheses in original)

As with AI, synthetic biology and genetic engineering are simultaneously presented as potential x-risks, but at the same time, *not* developing them is also a major risk to humanity:

> Synthetic biology is well-suited to address other cause areas. Climate change could be mitigated with biofuels, carbon capture, and sustainable production, or global health and development aided through improved access to food, clean water, and healthcare. Given the promise of synthetic biology, suboptimal development could represent permanent loss of great potential, constituting a p-risk[10]. What legal tools could help steer such technological progress? How could intellectual property law, economic development law such as taxes and subsidies (cf. Posner, 2008), trade law, and other legal fields influence development of the synthetic biology market?' (Winter et al., 2021, p. 76)

Instead, EAs advocate for 'differential development' (Beckstead, 2015), defined as slowing down 'dangerous' technologies while accelerating 'beneficial' technologies, particularly those that promise to 'ameliorate the hazards posed by other technologies' (Bostrom, 2002, section 9.4, para. 2). 'Differential development' means making sure that emerging technologies are developed in the *right* order and by people with the *correct* values, so as to avoid accidental or deliberate catastrophe (Wiblin, 2016). Again, this has the appearance of decelerating research and development, but EAs are in fact arguing in favor of developing technologies *they* designate as 'beneficial', assuming that their values are the correct ones, and that they possess knowledge of the right order of technological development. Ostensibly working to avoid what EAs call a negative 'values lock-in' (Karnofsky, 2021a; MacAskill, 2022; Tomasik, 2013), EAs desire to place the specific set of transformative technologies *they* believe are beneficial for humanity as global priorities. What, precisely, constitutes a beneficial technology (namely: *who* benefits?) is glossed as self-evident: to become an advanced, 'technologically mature', spacefaring civilization is assumed to be humanity's natural end goal. Once this assumption is taken for granted, what matters is ensuring that this occurs in a way that produces positive rather than negative value. This reasoning performs the necessary switch from pessimism about how terribly things could go wrong to optimism about steering humanity in a positive direction — it is a move wherein a process of envisioning a desirable future is intended to lead towards implementing plans and policies *now* to ostensibly help reach that future. The apocalyptic x-risk discourse creates a sense of urgency (requiring immediate action), alleviated through optimistic visioneering about steering humanity's trajectory toward a better future.

Notably, the future scenarios that EAs believe warrant serious consideration are restricted to those related to the futures EAs would/would not like to see, and to the technologies they hope to see developed: EAs urge careful consideration of one particular socio-technical future (which conveniently reflect the ideological perspective of the movement's founders and the investment portfolio of the

[9] Thus, to maintain defensive capabilities against malevolent actors we must pre-emptively facilitate R&D by those who can be trusted to be working on humanity's behalf. In this view, prohibitions on the development of synthetic biology, radical biotechnologies, or AGI would simply prevent well-intentioned actors from developing them, while malevolent actors would continue their efforts hidden from sight, and attain unstoppable power.

[10] These EAs are using 'p-risk' to signify a risk to humanity's *potential.*

movement's funders) and conclude that they must begin the process of steering global society toward that future. Treating these particular ideas about the future—which are by no means representative of the myriad futures that the rest of the planet's population might envision as good, desirable, or even technologically achievable (and thus worthy of investment)—amounts to tacitly endorsing them as somehow more real, more valid, than other peoples' visions. The silencing of alternative visions of the future facilitates *one* vision to become hegemonic (Nandy, 1996), relegating all other options to the realm of impossible or simply unthinkable — a kind of closure and stabilization, wherein one vision is naturalized as common-sense, and becomes the *only* one thinkable for a given period[11] (Bijker, 1997). The demographic homogeneity of the EAs (overwhelmingly white men of privileged backgrounds with a predilection for computer science, analytic philosophy, and technology) bespeaks a colonizing logic[12] wherein this particular group assumes the role of enlightened saviors to whom the future ought to be entrusted, lest the planet fall into the hands of (black/indigenous/female) Others (Gergan et al., 2020; Mitchell & Chaudhury, 2020) who could destroy humanity's 'potential' by derailing what EAs see as the *correct* global technological trajectory. Often speaking of 'civilization', EAs presume the 'Eurocentric universal', projecting specific Western worldviews and values onto humanity writ large (Ali, 2019). While sometimes careful to define 'civilization' as a synonym for all of humanity and future human-descended beings (Baum et al., 2019), in practice, the type of 'civilization' whose survival and flourishing is a precious duty to protect is a specifically Western-technologically-oriented society that can hardly be considered representative of all humanity. EAs perpetuate the gendered and racialized 'exclusionary hierarchy of humanity' (Gergan et al., 2020, p. 92) that mark related discourses of apocalyptic crisis, such as climate change and the Anthropocene (Gergan et al., 2020; Simpson, 2020; Whyte, 2018).

Crucially, these are not merely acts of visioning, but the elaboration of blueprints for the future that transhumanists/EAs desire to set in motion. By claiming to speak on behalf of the future, EAs and transhumanists authorize themselves 'to declare which of society's anxieties are misguided and what modes of governance stand in the way of the future' (Boenig-Liptsin & Hurlbut, 2016, p. 264) while also issuing directives regarding which *potential* events policymakers should ignore and which ones demand their attention (Mallard & Lakoff, 2011). Levitas (2013) argues that the articulation of a utopia (which she defines simply as 'the expression of the desire for a better way of being or of living', p. xii) is much more than the expression of a fantasy or even a goal; it is a *method* for working toward that goal (Levitas, 2013). She draws on the work of Ernst Bloch, who understood utopia as a form of anticipatory consciousness wherein 'the central idea of *not yet* carries the double sense of not *yet* (but expected, a future presence) and still *not* (a current absence and lack)' (Levitas, 2013, p. 6). Utopian visions do not merely predict or express hopes about a future, but performatively draft a roadmap, delineating the steps an actor believes are necessary to reach that future. The EAs' utopianism can be understood as not just the expression of their goal, but their method for steering the world toward that goal: it is an act of visioning that orients policymakers, shapes research agendas, and influences public debates (Grunwald, 2016). To control the image of the future is, in a sense, to delimit what kind of future is possible and thereby shape the future that actually comes into being (Shaw, 2021); efforts to 'shape the future' are, of course, merely interventions in the present that are expected to lead particular outcomes (Grunwald, 2019).

By carefully eliding fundamental questions of whether we collectively *want*, *should*, or even *can* create AGI and the other 'transformative' technologies, EA normalizes them as both inevitable and highly desirable; there is no choice regarding *whether* we should be trying to create AGI — it is happening

[11] In terms of being acceptable for the purpose of a research agenda, receiving funding, policy decisions, etc.

[12] For instance, even as they speak of catastrophic threats that would end the world as we know it, EAs overlook how the very same discourse of technological progress and Enlightenment that they celebrate spelled the apocalypse for many (indigenous) civilizations (Whyte, 2018). Relatedly, EAs breathlessly hype space colonization as humanity's 'cosmic endowment', treating 'colonization' as a marvelous frontier-pushing adventure, rather than the actual existential threat that 'colonization' has meant to the majority of the planet's population (Redfield, 2002).

whether we want it or not (they argue) — our only choice (in this view) is between ensuring safe AI that is aligned with human values, or allowing the creation of unsafe, uncontrolled, unaligned AI (Cremer & Kemp, 2021). EA flattens debate about future technological developments into binary terms: good AI vs. bad AI; survival and flourishing vs. collapse and extinction; technological maturity vs. technological stagnation. Framed in such Manichean terms, there is only one sensible path: to create safe, aligned AI and safe biotechnologies—and to do so as quickly as possible (albeit keeping in mind the need for balancing speed with safety, per Bostrom's 'differential development' clause), before more nefarious, ill-intentioned actors achieve their aims first. Speculative ethics are at work, transforming prognostications about hypothetical futures into concrete policy decisions and research agendas:

> When it comes to preparing for dangers that don't yet exist, such as transformative AI, I think it's incredibly valuable to prepare for things now that will have dramatic effects in the future, even if we're unclear about what those effects will be. We should get started on difficult problems now instead of leaving them for the next generation to tackle. (Cargill, in Jacobs, 2019)

The practical result is the shunting of young EAs into careers in technical AI research, AI policy work, biotechnology and biosecurity; a flood of funding to AI researchers and institutes promising that they are working to develop 'safe'/'aligned' AI and to researchers who similarly promise they are developing cutting edge biotechnologies to reduce suffering and benefit humanity; and the positioning of emerging technologies on global agendas as the most important political issue of the century.

6. Safety?

Across EA, the paeans to *safety* gild R&D and policy entrepreneurship for emerging technologies. It is beyond the scope of this paper to detail the scale of EA investments in emerging technology research and policymaking, but below are several examples of grants made by Open Philanthropy (OpenPhil, the main funder of the EA movement, which oversees Facebook co-founder Dustin Moskovitz's philanthropy) to illustrate how research that flies under the rubric of 'safety' or 'x-risk reduction' is still largely applied research.

Many of OpenPhil's grants made for biosecurity/pandemic preparedness go toward pre-emptively developing novel biotechnologies, in the name of ensuring they are developed *safely*. For instance, OpenPhil granted $4,748,881 for Kevin Esvelt's Sculpting Evolution lab at MIT (Open Philanthropy, 2019a, 2021b, 2022a, 2022c), $85,000 to Esvelt and Michael Specter to co-teach a course at MIT on 'longtermism' (Open Philanthropy, 2021a), and $5,318,000 to launch CEA's new Boston Biosecurity Hub, which hosts Esvelt's lab and other EAs working on biosecurity (Open Philanthropy, 2022b). OpenPhil granted $2,970,000 to Ed Boyden's synthetic neurobiology group at MIT, which works on brain mapping, i.e., the first step in creating digital minds (Open Philanthropy Project, 2016). Good Ventures (Moskovitz's foundation) has also committed $24,000,000 to the Arc Institute, a non-profit biotechnology research lab co-founded by Stripe billionaire Patrick Collison which aims 'to encourage cross-disciplinary biomedical innovation by removing bureaucratic hurdles posed by traditional funding structures' (Good Ventures, 2021). OpenPhil has also provided $8,346,600 to support the anti-aging research of Irina Conboy, who focuses on therapeutic blood exchange and rejuvenation through blood dilution techniques (Open Philanthropy, 2017, 2019b, 2023).

In AI, EAs position their own technical AI *alignment* research in opposition to AI *capabilities* research (the former is understood to pursue *safe* AI, and the latter is seen as having little-to-no concern for safety). But alignment and capabilities research advance the same project in a kind of good cop/bad cop routine, where both are pursuing the same ultimate goal — reinforcement learning, robotics, large language models (LLMs), etc., as a step towards the dream of creating AGI. It is far from clear that EA's investments in AI-safety and AI-alignment have produced 'safer' versions of these technologies; rather, it appears they accelerate (even if inadvertently) harms. The most popularly-discussed example

is EA's investment in OpenAI[13]: OpenPhil granted $30,000,000 to OpenAI in 2017 (Open Philanthropy Project, 2017); OpenAI's research team and board were populated by leading EAs; and OpenAI was hyped as EA's favored 'AI charity' (FTX, 2020; Piper, 2019). Yet far from slowing down or ensuring the safe development of AI, OpenAI's ChatGPT (which generates text that appears to have been created by a human) and DALL-E (a text-to-image AI system that produces pictures) initiated an industry-wide race to build larger models whose dangers include 'creating child pornography, perpetuating bias, reinforcing stereotypes, and spreading disinformation en masse' (Gebru, 2022).

To give another less-discussed example: OpenPhil provided generous support to Pieter Abbeel under the rubric of AI safety[14] through his positions at two OpenPhil-funded organizations—OpenAI and the Berkeley Center for Human-Compatible AI (CHAI); in 2020, Abbeel cofounded Covariant AI (Covariant, 2020). Far from serving a grand humanitarian effort, Covariant makes robotic warehouse pickers, which replace human workers for online retailers (Hao, 2020; Knight, 2020; Vincent, 2020) — a sector of labor beset by strikes and efforts to unionize for better pay and safer working conditions (Sainato, 2019). Regardless of their ethical *intentions*, researchers like Abbeel aren't protecting humanity from unaligned AI, they are protecting the interests of the capitalist class in the most predictable way: by replacing unruly human workers with machines (Berg, 1980; MacKenzie, 1984; Winner, 1980).

7. Conclusions

This article aimed to show how the transhumanists linked their project to humanitarian efforts, i.e., protecting the vulnerable, saving lives, preventing suffering and harm, and ensuring a flourishing future for all. Whereas *life-saving transhumanitarianism* was oriented around techno-optimism, *x-risk transhumanitarianism* focused on risks and dangers, issuing sober calls for the *safe* development of emerging technologies, and the protection of future generations from potential harms. By cultivating the EA movement and embedding *x-risk transhumanitarianism* within it—now under the banner of 'longtermism'—the transhumanists used EA as a Trojan horse to elevate their vision for humanity's future onto mainstream global policy and research agendas.[15]

While the alarmist discourse of human extinction may seem counterintuitive as an advocacy technique, in fact, it has proven enormously effective. First, it *inoculates* the transhumanists and EAs from the charge of unbridled techno-optimism, since they openly admit that emerging technologies pose great risks and warrant public concern, and even stake themselves as the vanguard identifying and mitigating such risks. As Barthes observed, portraying the drawbacks of an idea or program is 'a paradoxical but incontrovertible means of exalting it' (Barthes, 1972 [1957], p. 40). EAs position themselves as so utterly concerned about the potential negative effects of technology that they, themselves, are spearheading research on AI to ensure safety and prevent harm; acting simultaneously as AI's biggest champions *and* doomsayers, EAs are positioned to frame global discussion and set the terms of debate, restricting critique to that which ultimately serves transhumanist (and industry) goals. It comes as little surprise that EA has been taken up enthusiastically by Silicon Valley elites heavily invested in AI, who can present themselves as deeply concerned about AI risks, while projecting that risk on to hypothetical

[13] Space constraints prevent further elaboration of the extent of EA investments in AI-safety/AI-alignment in the present article; suffice it to say that EA funders like OpenPhil, Jaan Tallinn, and the now-defunct FTX have invested hundreds of millions of dollars into organizations led and/or populated by EAs working on AI/x-risk, including: Anthropic, Alignment Research Center, Ought, Redwood Research, Center for Human-Compatible AI (CHAI) at UC Berkeley, the Center for the Governance of AI, Cooperative AI Foundation, AI Impacts, AI Objectives Research Institute, the Center for AI Safety, Hofvarpnir Studies, Aligned AI, Center for the Study of Existential Risk (CSER), Center for Security and Emerging Technologies (CSET).

[14] OpenPhil provided $5.5 million to launch CHAI, where Abbeel was an affiliated researcher; Abbeel and his colleague were awarded a joint grant of $1,145,000 from OpenPhil in 2018 (Open Philanthropy Project, 2018); and Abbeel was among the highest paid employees of OpenAI at the time when its funding was primarily the OpenPhil grant (OpenAI, 2017, 2018).

[15] Crucially, EAs can always fall back on public-facing EA discourse (focused on saving lives in poor countries) to protect EA's reputation for 'doing good better' when confronted with criticism of the AI/x-risk agenda.

future scenarios — a convenient distraction from the known, *actually-existing* problems that current machine learning technologies produce and reinforce (e.g., algorithmic bias, misinformation, automation, privacy concerns, etc.).

More importantly, the focus on threats/risks *performatively* inscribes the transhumanists' desired technological research agenda and vision in the sociotechnical imaginary as both imminent and inevitable. The vivid articulation of a fear conjures the thing-to-be-feared into existence. Just as the mythical 'missile gap' drove the arms race during the Cold War and thereby help *manifest* the very technological threat it was intended to ameliorate (Amadae, 2003; Ellsberg, 2017), by outlining a comprehensive and concrete research program that ostensibly *responds to* the (potential) dangers of transhumanism's desired technological developments, such technologies become real (even if they do not exist). EA's x-risk and AI-safety programs are packaged and presented as responses to what could go *wrong*; but by defining what could go 'wrong', a specific vision of what it means for things to go 'right' is smuggled in through the back door as a taken-for-granted assumption. Like a film negative producing an image, the very same program that ostensibly addresses the *problems* with a particular vision for the future simultaneously produces that vision's contours. To borrow a metaphor from anthropologist Annelise Riles (2001), it is like creating a drawing of a figure by sketching not the figure itself, but the air that surrounds it.

References

80000 Hours. (2021). *Our current view of the world's most pressing problems [archived 10/26/2021]*. https://archive.md/x7GK1

Achenbach, J. (2015, December 27). The AI Anxiety: Big-name scientists worry that runaway artificial intelligence could pose a threat to humanity. *Washington Post*. http://www.washingtonpost.com/sf/national/2015/12/27/aianxiety

Ali, S. M. (2019). "White Crisis" And/As "Existential Risk", or the Entangled Apocalypticism of Artificial Intelligence. *Zygon®*, *54*(1), 207-224.

Amadae, S. M. (2003). *Rationalizing Capitalist Democracy: The Cold War Origins of Rational Choice Liberalism*. University of Chicago Press.

Andersen, R., & Bostrom, N. (2012, March 6). We're Underestimating the Risk of Human Extinction (Interview with Nick Bostrom). *The Atlantic*. http://www.theatlantic.com/technology/archive/2012/03/were-underestimating-the-risk-of-human-extinction/253821/

Anderson, B. (2007). Hope for nanotechnology: anticipatory knowledge and the governance of affect. *Area*, *39*(2), 156-165.

Barbrook, R., & Cameron, A. (1996). The Californian Ideology.

Barrow, J. D., & Tipler, F. J. (1986). *The Anthropic Cosmological Principle*. Clarendon.

Barthes, R. (1972 [1957]). *Mythologies* (A. Lavers, Trans.). The Noonday Press.

Baum, S., Armstrong, S., Ekenstedt, T., Häggström, O., Hanson, R., Kuhlemann, K., Maas, M., MIller, J. D., Salmela, M., Sandberg, A., Sotala, K., Torres, P., Turchin, A., & Yampolskiy, R. (2019). Long-Term Trajectories of Human Civilization. *Foresight*, *21*(1), 53-83.

Beckstead, N. (2013). *On the Overwhelming Importance of Shaping the Far Future* [PhD, Philosophy, Rutgers University]. New Brunswick.

Beckstead, N. (2015). Differential technological development: Some early thinking. *GiveWell blog*. https://web.archive.org/web/20191224091520/https://blog.givewell.org/2015/09/30/differential-technological-development-some-early-thinking/

Bell, T. W., & O'Connor, M. T. (1988). Introduction. *Extropy*, *1*(Fall 1988), 1-13. https://archive.org/details/extropy-01/mode/2up

Berg, M. (1980). *The Machinery Question and the Making of Political Economy, 1815-1848*. Cambridge University Press.

Bijker, W. E. (1997). *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. MIT Press.

Boenig-Liptsin, M., & Hurlbut, J. B. (2016). Technologies of Transcendence at Singularity University. In J. B. Hurlbut & H. Tirosh-Samuelson (Eds.), *Perfecting Human Futures: Transhumanist Visions and Technological Imaginations* (pp. 239-268). Springer.

Bostrom, N. (1999). *The Transhumanist FAQ - Version of May 13, 1999 [archived 8/17/2000]*. https://web.archive.org/web/20000817094531/http://www.transhumanist.org/

Bostrom, N. (2001). *Nick Bostrom's Home Page*. https://web.archive.org/web/20010203160400/http://www.nickbostrom.com/

Bostrom, N. (2002). Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards. *Journal of Evolution and Technology*, *9*. https://www.nickbostrom.com/existential/risks.html

Bostrom, N. (2003a). Astronomical Waste. https://www.nickbostrom.com/astronomical/waste.pdf

Bostrom, N. (2003b). Ethical Issues in Advanced Artificial Intelligence. https://www.nickbostrom.com/ethics/ai.html

Bostrom, N. (2003c). When machines outsmart humans. *Futures*, *35*(7).

Bostrom, N. (2005a). The Fable of the Dragon Tyrant. *Journal of Medical Ethics*, *31*, 273-277.

Bostrom, N. (2005b). A History of Transhumanist Thought. *Journal of Evolution and Technology*, *14*(April).

Bostrom, N. (2006). Letter from Utopia. https://web.archive.org/web/20061115101128/http://www.nickbostrom.com/utopia.html

Bostrom, N. (2008a). Where Are They? Why I hope the search for extraterrestrial life finds nothing. *MIT Technology Review*. https://www.technologyreview.com/2008/04/22/220999/where-are-they/

Bostrom, N. (2008b). Why I Want to Be a Posthuman When I Grow Up. https://www.nickbostrom.com/posthuman.pdf
Bostrom, N. (2011a). *The Concept of Existential Risk [archived 10/05/2011]*. Future of Humanity Institute. https://web.archive.org/web/20111005100617/http://www.existential-risk.org:80/concept.pdf
Bostrom, N. (2011b). In Defense of Posthuman Dignity. In G. R. Hansell & W. Grassie (Eds.), *H±: Transhumanism and Its Critics* (pp. 55-66). Metanexus.
Bostrom, N. (2013). Existential Risk Prevention as Global Priority. *Global Policy*, *4*(1), 15-31.
Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
Bostrom, N., & Ćirković, M. (Eds.). (2008). *Global Catastrophic Risks*. Oxford University Press.
Bostrom, N., Dafoe, A., & Flynn, C. (2016). Public Policy and Superintelligent AI: A Vector Field Approach.
Bostrom, N., & High, P. (2016). Nick Bostrom On The Single Most Important Challenge That Humanity Has Ever Faced. *Forbes*. https://www.forbes.com/sites/peterhigh/2016/06/27/nick-bostrom-on-the-single-most-important-challenge-that-humanity-has-ever-faced/
Bostrom, N., & Ord, T. (2006). The Reversal Test: Eliminating Status Quo Bias in Applied Ethics. *Ethics*, *116*, 656-679.
Burdett, M. S. (2015). The Religion of Technology: Transhumanism and the Myth of Progress. In C. Mercer & T. J. T. Rothen (Eds.), *Religion and Transhumanism: The Unknown Future of Human Enhancement* (pp. 131-148). Praeger.
Centre for Effective Altruism. (2018, May). *CEA's Current Thinking [archived 11/29/2018]*. https://web.archive.org/web/20181129224534/https://www.centreforeffectivealtruism.org/ceas-current-thinking/
Centre for Effective Altruism. (2020, July 18). Long Term Future Fund [archived 7/18/2020]. *Effective Altruism Funds*. http://archive.is/hF9eh
Collison, P., & Cowen, T. (2019, July 30). We Need a New Science of Progress: Humanity needs to get better at knowing how to get better. *The Atlantic*. https://www.theatlantic.com/science/archive/2019/07/we-need-new-science-progress/594946/
Covariant. (2020). Q&A With the Founders. *Medium* https://medium.com/covariant-ai/q-a-with-the-founders-dbd41dfcb208
Cowen, T. (2010). *The Great Stagnation*. Dutton.
Cowen, T. (2013). *Average is over: Powering America beyond the great stagnation*. Dutton.
Cremer, C. Z., & Kemp, L. (2021). Democratising Risk: In Search of a Methodology to Study Existential Risk. https://ssrn.com/abstract=3995225
de Grey, A. (2006a). Why should you do whatever you can to expedite the defeat of human aging? *Methuselah Foundation/M-prize*. https://web.archive.org/web/20070310222832/http://www.mprize.org/index.php?pagename=whyaging
de Grey, A. (2006b). Why we should do all we can to hasten the defeat of human ageing. *SENS*. https://web.archive.org/web/20070206024720/http://www.sens.org/concerns.htm
de Grey, A. (2007, December 20). Old People Are People Too: Why It Is Our Duty to Fight Aging to the Death. *Cato Unbound*. https://www.cato-unbound.org/2007/12/03/aubrey-de-grey/old-people-are-people-too-why-it-our-duty-fight-aging-death
Dickens, M. (2020). "Disappointing Futures" Might Be As Important as Existential Risks. *EA Forum*. https://forum.effectivealtruism.org/posts/9AYmbh25eKLojeQGe/disappointing-futures-might-be-as-important-as-existential
EA Forum. (2021). *Tag - Longtermism [archived 4/6/2021]*. https://web.archive.org/web/20210406105804/https://forum.effectivealtruism.org/tag/longtermism
EA forum. (2022). *Tag - Hinge of History [archived 6/8/2022]*. https://web.archive.org/web/20220608175228/https://forum.effectivealtruism.org/topics/hinge-of-history
EA Forum. (n.d.). Trajectory Change. *EA Forum*. https://forum.effectivealtruism.org/tag/trajectory-change
Ellsberg, D. (2017). *The Doomsday Machine: Confessions of a Nuclear War Planner*. Bloomsbury.

Extropy Institute. (2003). *Extropy: Journal of Transhumanist Solutions - Home [archived 5/27/2005]*. https://web.archive.org/web/20050527011103/http://spock.extropy.org/ideas/journal/
Fisher, R. (2020, September 24). Are We Living at the Hinge of History? *BBC Future*. https://www.bbc.com/future/article/20200923-the-hinge-of-history-long-termism-and-existential-risk
Forethought Foundation. (2018). *Research Areas - Overview [archived 11/25/2018]*. https://web.archive.org/web/20181125024907/https://www.forethought.org/research-overview/
FTX. (2020). *FTX - About*. Retrieved June 2 from https://web.archive.org/web/20200602075639/https://about.ftx.com/
FTX Future Fund. (2022a). *Areas of Interest*. https://web.archive.org/web/20220321205921/https://ftxfuturefund.org/area-of-interest/
FTX Future Fund. (2022b). *Project Ideas*. https://web.archive.org/web/20220321205440/https://ftxfuturefund.org/projects/
Gebru, T. (2022). Effective Altruism is Pushing a Dangerous Brand of 'AI Safety'. *Wired*. https://www.wired.com/story/effective-altruism-artificial-intelligence-sam-bankman-fried/
Geraci, R. (2010). *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. Oxford University Press.
Gergan, M., Smith, S., & Vasudevan, P. (2020). Earth beyond repair: Race and apocalypse in collective imagination. *Environment and Planning D: Society and Space*, *38*(1), 91-110.
Gleiberman, M. (2023). *Effective Altruism and the Strategic Ambiguity of 'Doing Good'*. (IOB Discussion Paper, 2023.01). Institute of Development Policy (IOB), University of Antwerp. https://www.uantwerpen.be/en/research-groups/iob/publications/discussion-papers/dp-2023/
Good Ventures. (2021). *Grant - Arc Institute - general support ($24,000,000)*. https://www.goodventures.org/our-portfolio/grants/arc-institute-general-support
Greaves, H., MacAskill, W., O'Keeffe-O'Donovan, R., & Trammell, P. (2019). *A Research Agenda for the Global Priorities Institute*. Global Priorities Institute, University of Oxford. https://globalprioritiesinstitute.org/wp-content/uploads/gpi-research-agenda.pdf
Grunwald, A. (2016). What Does the Debate on (Post)human Futures Tell Us?: Methodology of Hermeneutical Analysis and Vision Assessment. In J. B. Hurlbut & H. Tirosh-Samuelson (Eds.), *Perfecting Human Futures: Transhuman Visions and Technological Imaginations* (pp. 35-50). Springer.
Grunwald, A. (2019). Shaping the Present by Creating and Reflecting Futures. In A. Lösch, A. Grunwald, M. Meister, & I. Schulz-Schaeffer (Eds.), *Socio-Technical Futures Shaping the Present: Empirical Examples and Analytical Challenges* (pp. 17-36). Springer.
Hall, M. (2017). *The Bioethics of Enhancement: Transhumanism, Disability, and Biopolitics*. Lexington Books.
Halstead, J. (2019). *Existential Risk: Cause Area Report [archived 5/24/2020]*. Founders Pledge. https://web.archive.org/web/20200524091707/https://founderspledge.com/research/fp-existential-risk
Hao, K. (2020). AI-powered robot warehouse pickers are now ready to go to work. *MIT Technology Review*. https://www.technologyreview.com/s/615109/ai-powered-robot-warehouse-pickers-are-now-ready-to-go-to-work/
Harrington, M. (2022, July 23). Peter Thiel on the dangers of progress. *UnHerd*. https://unherd.com/2022/07/peter-thiel-on-the-dangers-of-progress/
Harris, J. (2009). Enhancements are a Moral Obligation. In J. Savulescu & N. Bostrom (Eds.), *Human Enhancement* (pp. 131-154). Oxford University Press.
Harris, J. (2019). How tractable is changing the course of history? *EA Forum*. https://forum.effectivealtruism.org/posts/QrGeEGMpKMEMsdThL/how-tractable-is-changing-the-course-of-history
Hauskeller, M. (2016). *Mythologies of Transhumanism*. Palgrave MacMillan.
Hughes, J. J. (2002). The Politics of Transhumanism (Version 2.0, March 2002). *ChangeSurfer*. http://changesurfer.com/Acad/TranshumPolitics.htm
Humphrey, S. (2004). No death, please, I'm bionic. *NOW Toronto*. https://nowtoronto.com/news/no-death-please-im-bionic

Jacobs, A. J. (2019). The Billionaire and the (Very, Very) Far Future (interview with Natalie Cargill and Ben Delo). https://www.linkedin.com/pulse/billionaire-very-far-future-a-j-jacobs/
John, T. M. (2019). Institutions for Future Generations. *EA Forum*. https://forum.effectivealtruism.org/posts/op93xvHkJ5KvCrKaj/institutions-for-future-generations
Jones, R. (2009). The Economy of Promises. *Nature Nanotechnology*, *3*, 65. http://www.softmachines.org/wordpress/?p=449
Karnofsky, H. (2021a, July 14). All Possible Views About Humanity's Future Are Wild. *Cold Takes*. https://www.cold-takes.com/all-possible-views-about-humanitys-future-are-wild/
Karnofsky, H. (2021b). The "most important century" blog post series. *Cold Takes*. https://archive.is/GPwbx
Knight, W. (2020). AI Helps Warehouse Robots Pick Up New Tricks. *Wired*. https://www.wired.com/story/ai-helps-warehouse-bots-pick-new-skills/
Koch, N. (2021). Whose apocalypse? Biosphere 2 and the spectacle of settler science in the desert. *Geoforum*, *124*, 36-45.
Levitas, R. (2013). *Utopia as Method: The Imaginary Reconstitution of Society*. Palgrave MacMillan.
MacAskill, W. (2015). *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. Gotham Books.
MacAskill, W. (2019). 'Longtermism' [archived 7/19/2020]. *EA Forum*. https://web.archive.org/web/20200719203240/https://forum.effectivealtruism.org/posts/qZyshHCNkjs3TvSem/longtermism
MacAskill, W. (2022). *What We Owe the Future*. Basic Books.
MacAskill, W., & Islam, H. (2020). Q&A with Will MacAskill (transcript of interview at EAGxVirtual 2020). *EA Forum*. https://forum.effectivealtruism.org/posts/sFj7EstDYacf6GJWF/q-and-a-with-will-macaskill
Macaskill, W., & Tarsney, C. (2019). *Topics in Global Priorities Research (Trinity term 2019)*. https://globalprioritiesinstitute.org/topics-in-global-priorities-research/
MacKenzie, D. (1984). Marx and the Machine. *Technology and Culture*, *25*(3), 473-502.
Mackinac Center. (n.d.). *The Overton Window*. https://www.mackinac.org/OvertonWindow
Mallard, G., & Lakoff, A. (2011). How claims to know the future are used to understand the present. In C. Camic, N. Gross, & M. Lamont (Eds.), *Social Knowledge in the Making* (pp. 339-377). University of Chicago Press.
Matheny, J. G. (2006). Reducing the risk of human extinction. *Accelerating Future*. https://web.archive.org/web/20061024143031/http://www.acceleratingfuture.com/papers/extinction.htm
Miller, P., & James, W. (2006). Stronger, longer, smarter, faster. In P. Miller & J. Wilsdon (Eds.), *Better Humans? The Politics of Human Enhancement and Life Extension* (pp. 13-28). Demos.
Mitchell, A., & Chaudhury, A. (2020). Worlding beyond 'the''end'of 'the world': White apocalyptic visions and BIPOC futurisms. *International Relations*, *34*(3), 309-332.
Moravec, H. (1988). *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.
Mosco, V. (2005). *The Digital Sublime: Myth, Power, and Cyberspace*. MIT Press.
Nandy, A. (1996). Bearing witness to the future. *Futures*, *28*(6-7), 636-639.
Ngo, R. (2020). Thiel on Progress and Stagnation. *LessWrong*. https://www.lesswrong.com/posts/Xqcorq5EyJBpZcCrN/thiel-on-progress-and-stagnation
Nordmann, A. (2007). If and Then: A Critique of Speculative NanoEthics. *Nanoethics*, *1*, 31-46.
Nordmann, A., & Rip, A. (2009). Mind the Gap Revisited. *Nature - Nanotechnology*, *4*, 273-274.
Nuland, S. (2005, February 1). Do You Want to Live Forever? *MIT Technology Review*. https://www.technologyreview.com/2005/02/01/231686/do-you-want-to-live-forever/
Open Philanthropy. (2017). *Grant - UC Berkeley - Aging Research (Irina Conboy) ($5,000,000)*. https://web.archive.org/web/20200925104555/https://www.openphilanthropy.org/focus/scientific-research/uc-berkeley-aging-related-research-conboy-2017
Open Philanthropy. (2019a). *Grant - MIT Media Lab - Kevin Esvelt's Research ($1,000,000)*. https://www.openphilanthropy.org/grants/massachusetts-institute-of-technology-media-lab-kevin-esvelts-research-2019/

Open Philanthropy. (2019b). *Grant - UC Berkeley - Aging Research (Irina Conboy) ($304,000)*. https://web.archive.org/web/20200924003535/https://www.openphilanthropy.org/focus/scientific-research/miscellaneous/uc-berkeley-aging-research-irina-conboy-2019
Open Philanthropy. (2021a). *Grant - Course development support: Michael Specter co-teaching with Kevin Esvelt ($85,000)*. https://www.openphilanthropy.org/grants/michael-specter-course-development-support/
Open Philanthropy. (2021b). *Grant - MIT Media Lab - Kevin Esvelt's Research ($1,000,000)*. https://www.openphilanthropy.org/grants/massachusetts-institute-of-technology-media-lab-kevin-esvelts-research-2021/
Open Philanthropy. (2022a). *Grant - Berkeley Existential Risk Initiative (BERI) to support Kevin Esvelt's Sculpting Evolution group at MIT Media Lab ($100,000)*. https://www.openphilanthropy.org/grants/berkeley-existential-risk-initiative-support-for-kevin-esvelts-research/
Open Philanthropy. (2022b). *Grant - Centre for Effective Altruism, Biosecurity Coworking Space ($5,318,000)*. https://web.archive.org/web/20221216132005/https://www.openphilanthropy.org/grants/centre-for-effective-altruism-biosecurity-coworking-space/
Open Philanthropy. (2022c). *Grant - MIT Media Lab - Biosecurity Research led by Kevin Esvelt ($2,648,881)*. https://web.archive.org/web/20230327103232/https://www.openphilanthropy.org/grants/massachusetts-institute-of-technology-media-lab-biosecurity-research-2022/
Open Philanthropy. (2023). *Grant - UC Berkeley, Aging Research - Irina Conboy ($3,042,600)*. https://www.openphilanthropy.org/grants/university-of-california-berkeley-aging-research-irina-conboy-2023/
Open Philanthropy Project. (2016). *Grant - MIT Synthetic Neurobiology Group (Ed Boyden) ($2,970,000)*. https://web.archive.org/web/20181128192244/https://www.openphilanthropy.org/focus/scientific-research/miscellaneous/massachusetts-institute-technology-synthetic-neurobiology-group
Open Philanthropy Project. (2017). *Grant - OpenAI - General Support* [Grant]. https://web.archive.org/web/20170502125156/http://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/openai-general-support
Open Philanthropy Project. (2018). *Grant - UC Berkeley AI safety research ($1,145,000)*. https://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/university-of-california-berkeley-artificial-intelligence-safety-research-2018
OpenAI. (2017). *Return of Organization Exempt From Income Tax Form 990 - 2016*.
OpenAI. (2018). *Return of Organization Exempt From Income Tax - Form 990, 2017*.
Ord, T. (2020). *The Precipice*. Hachette Books.
Parfit, D. (1984). *Reasons and Persons*. Clarendon.
Pasek, J. (1992). Obligations to Future Generations: A Philosophical Note. *World Development*, *20*(4), 513-521.
Pearce, D. (2007). The Abolitionist Project. *Abolitionist*. https://www.abolitionist.com/
Pearce, D. (2010). Top Five Reasons Transhumanism Can Eliminate Suffering. *h+ Magazine*. https://web.archive.org/web/20101025003628/http://www.hplusmagazine.com/editors-blog/top-five-reasons-transhumanism-can-eliminate-suffering
Petrucci, M. (1998). Future Generations: A right way forward? *Soundings*(9), 42-55.
Pinker, S. (2015). Human Progress Quantified. *Edge*. https://www.edge.org/annual-question/2016/response/26616
Pinto, A. T. (2019). Capitalism with a Transhuman Face. *Third Text*, 1-22. https://doi.org/10.1080/09528822.2019.1625638
Piper, K. (2019, April 17). Why the world's leading AI charity decided to take billions from investors. *Vox - Future Perfect*. https://www.vox.com/future-perfect/2019/4/17/18301070/openai-greg-brockman-ilya-sutskever
Ranisch, R., & Lorenz Sorgner, S. (2014). Introducing Post- and Transhumanism. In R. Ranisch & S. Lorenz Sorgner (Eds.), *Post- and Transhumanism* (pp. 7-27). Peter Lang.

Redfield, P. (2002). The half-life of empire in outer space. *Social Studies of Science*, *32*(5-6), 791-825.
Rees, G. (2018). Lunch With: Nick Bostrom. *Gareth Rees*. http://gdlrees.co.uk/writing/nickbostrom
Riles, A. (2001). *The Network Inside Out*. University of Michigan Press.
Sainato, M. (2019). 'We are not robots': Amazon warehouse employees push to unionize. *The guardian*. https://www.theguardian.com/technology/2019/jan/01/amazon-fulfillment-center-warehouse-employees-union-new-york-minnesota
Sandberg, A. (2015). Transhumanism and the Meaning of Life. In C. Mercer & T. J. Trothen (Eds.), *Religion and Transhumanism: The Unkonw Future of Human Enhancement* (pp. 3-22). Praeger.
Savulescu, J., & Sandberg, A. (2008). Neuroenhancement of Love and Marriage: The Chemicals Between Us. *Neuroethics*, *1*, 31-44.
Schuster, J., & Woods, D. (2021). *Calamity Theory: Three Critiques of Existential Risk*. University of Minnesota Press.
Shaw, M. (2021). Billionaire capitalists are designing humanity's future. Don't let them. *The guardian*. https://www.theguardian.com/commentisfree/2021/feb/05/jeff-bezos-elon-musk-spacex-blue-origin
Simpson, M. (2020). The Anthropocene as colonial discourse. *Environment and Planning D: Society and Space*, *38*(1), 53-71. https://doi.org/10.1177/0263775818764
Singularity Institute for Artificial Intelligence. (2002a). *Six Reasons Why Small Donations Matter*. https://web.archive.org/web/20021220084208/http://singinst.org/small-donations-matter.html
Singularity Institute for Artificial Intelligence. (2002b). Why Work Toward the Singularity? https://web.archive.org/web/20021208075615/http://singinst.org/why-singularity.html
Smart, J. J. C. (1973). An Outline of a System of Utilitarian Ethics. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and Against*. Cambridge University Press.
Soares, N. (2018). 2018 Update: Our New Research Directions. *MIRI*. https://intelligence.org/2018/11/22/2018-update-our-new-research-directions/
Swyngedouw, E. (2013). Apocalypse Now! Fear and Doomsday Pleasures. *Capitalism Nature Socialism*, *24*(1), 9-18. https://doi.org/10.1080/10455752.2012.759252
Thiel, P. (2013). *Keynote - Effective Altruism Summit, 2013*. https://youtu.be/h8KkXcBwHec
Thiel, P., & Cowen, T. (2015). The Future of Innovation: A conversation between Tyler Cowen and Peter Thiel. *Conversations with Tyler*. https://youtu.be/i_yJTCDU4uE
Thiel, P., & Masters, B. (2014). *Zero to One: Notes on Startups, or, How to Build the Future*. Virgin/Ebury.
Tirosh-Samuelson, H. (2011). Engaging Transhumanism. In G. R. Hansell & W. Grassie (Eds.), *H±: Transhumanism and Its Critics* (pp. 19-52). Metanexus.
Tirosh-Samuelson, H. (2014). Religion. In R. Ranisch & S. Lorenz Sorgner (Eds.), *Post- and Transhumanism: An Introduction* (pp. 49-71). Peter Lang.
Tirosh-Samuelson, H., & Hurlbut, J. B. (2016). Introduction: Technology, Utopianism and Eschatology. In J. B. Hurlbut & H. Tirosh-Samuelson (Eds.), *Perfecting Human Futures: Transhuman Visions and Technological Imaginations* (pp. 1-32). Springer.
Todd, B. (2013). How Important are Future Generations? *80,000 Hours*. https://web.archive.org/web/20130829023754/http://80000hours.org/blog/245-how-important-are-future-generations
Todd, B. (2017). If You Want to Do Good, Here's Why You Should Focus on Future Generations [archived: 10/25/2017]. *80,000 Hours*. https://web.archive.org/web/20171025030128/https://80000hours.org/articles/future-generations/
Todd, B. (2018). Presenting the long-term value thesis [archived 8/8/2018]. *80,000 Hours*. https://web.archive.org/web/20180808135306/https://80000hours.org/articles/future-generations/
Todd, B. (2019, April 22). Introducing longtermism [archived: 10/19/2019; note that the article remains dated 2017]. *80,000 Hours*. https://web.archive.org/web/20191019032811/https://80000hours.org/articles/future-generations/

Todd, B. (2020). Why I've come to think global priorities research is more important than I thought. *80,000 Hours*. https://80000hours.org/2020/08/global-priorities-research-update/

Todd, B., & 80000 Hours Team. (2019). *A guide to using your career to help solve the world's most pressing problems [archive, 12/29/2019]*. 80,000 Hours. https://web.archive.org/web/20191229205609/https://80000hours.org/key-ideas/

Tomasik, B. (2013). The haste consideration, revisited. https://felicifia.github.io/thread/824.html

Verdoux, P. (2009). Transhumanism, Progress and the Future. *Journal of Evolution and Technology*, *20*(2), 49-69.

Vincent, J. (2020, January 29). AI-powered robot pickers will be the next big work revolution in warehouses. *The Verge*. https://www.theverge.com/2020/1/29/21083313/robot-picking-warehouses-logistics-ai-covariant-stealth

Walker, M. (2007). Superlongevity and Utilitarianism. *Australasian Journal of Philosophy*, *85*(4), 581-595.

Weinstein, E., & Thiel, P. (2019). Episode #1: An Era of Stagnation & Universal Institutional Failure. *The Portal*. https://youtu.be/nM9f0W2KD5s

Whittlestone, J. (2017, November 16). *The Long Term Future*. Retrieved April 28 from https://web.archive.org/web/20200220215336/https://www.effectivealtruism.org/articles/cause-profile-long-run-future/

Whyte, K. (2018). Indigenous science (fiction) for the Anthropocene: Ancestral dystopias and fantasies of climate change crises. *Environment and Planning E: Nature and Space*, *1*(1-2), 224-242.

Whyte, K. (2021). Against Crisis Epistemology. In B. Hokowhitu, A. Moreton-Robinson, L. Tuhiwai Smith, C. Andersen, & S. Larkin (Eds.), *Routledge Handbook of Critical Indigenous Studies* (pp. 52-64). Routledge.

Wiblin, R. (2016). Transcript: Making sense of long-term indirect effects - Rob Wiblin's August 2016 EA Global Talk *jefftk*. https://www.jefftk.com/p/transcript-making-sense-of-long-term-indirect-effects

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1 Modern Technology: Problem or Opportunity, Winter 1980), 121-136.

Winter, C., Schuett, J., Martinez, E., Van Arsdale, S., Araujo, R., Hollman, N., Sebo, J., Stawasz, A., O'Keefe, C., & Rotola, G. (2021). Legal Priorities Research: A Research Agenda. https://www.legalpriorities.org/research_agenda.pdf

Zuckerberg, M., Collison, P., & Cowen, T. (2019). A Conversation with Mark Zuckerberg, Patrick Collison and Tyler Cowen. *Meta*. https://about.fb.com/news/2019/11/a-conversation-with-mark-zuckerberg-patrick-collison-and-tyler-cowen/

University of Antwerp
IOB | Institute of
Development Policy