f i ® s t m ¤ ñ d @ ¥
PEER-REVIEWED JOURNAL ON THE INTERNET

The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence

by Timnit Gebru and Émile P. Torres

Abstract

The stated goal of many organizations in the field of artificial intelligence (AI) is to develop artificial general intelligence (AGI), an imagined system with more intelligence than anything we have ever seen. Without seriously questioning whether such a system can and should be built, researchers are working to create “safe AGI” that is “beneficial for all of humanity.” We argue that, unlike systems with specific applications which can be evaluated following standard engineering principles, undefined systems like “AGI” cannot be appropriately tested for safety. Why, then, is building AGI often framed as an unquestioned goal in the field of AI? In this paper, we argue that the normative framework that motivates much of this goal is rooted in the Anglo-American eugenics tradition of the twentieth century. As a result, many of the very same discriminatory attitudes that animated eugenicists in the past (*e.g.*, racism, xenophobia, classism, ableism, and sexism) remain widespread within the movement to build AGI, resulting in systems that harm marginalized groups and centralize power, while using the language of “safety” and “benefiting humanity” to evade accountability. We conclude by urging researchers to work on defined tasks for which we can develop safety protocols, rather than attempting to build a presumably all-knowing system such as AGI.

Contents

1. Introduction

Recent years have seen a resurgence of the goal to build artificial general intelligence (AGI), a system defined differently by the various people and organizations that seek to build it. For instance, OpenAI defines AGI as “highly autonomous systems that outperform humans at most economically valuable work” [1]. Pennachin and Goertzel, who popularized the term in 2007, define it as “a software program that can solve a variety of complex problems in a variety of different domains, and that controls itself autonomously, with its own thoughts, worries, feelings, strengths, weaknesses and predispositions" [2]. Peter Voss, who claims to have helped coin the term, defines it as "a computer system that matches or exceeds the real time cognitive (not physical) abilities of a smart, well-educated human" [3]. And prominent AI researchers Stuart Russell and Peter Norvig define it as "a universal algorithm for learning and acting in any environment" [4]. While a number of the researchers who coined the term "artificial intelligence" (AI) in 1955 had the goal of creating "machines performing the most advanced human thought activities" (McCarthy, *et al.*, 1955), this goal was abandoned by many scholars in the field by the 1990s, in part because they did not want to be associated with grandiose claims that researchers didn't deliver on [5]. Such claims led to the "AI winters" of the 1970s and 1990s, with much of the research focused on building "general intelligence" losing funding. After that, the field was mostly focused on building specialized systems that some call "narrow AI" [6].

Recently, however, there has been a proliferation of organizations aiming to build AGI and asserting that their products are close to achieving it (Bubeck, *et al.*, 2023; Cuthbertson, 2022). While a number of researchers have debated whether or not various methodologies can achieve AGI (Shevlin, *et al.*, 2019; Agüera y Arcas and Norvig, 2023; Silver, *et al.*, 2021), we have seen little discussion of why AGI is considered desirable by many in the field of AI, and whether this is a goal that should be pursued — or is even possible in the first place. The quest to build what seems like an all-knowing system capable of performing any task under any circumstance has already resulted in many documented harms to marginalized groups, including worker exploitation (Gray and Suri, 2019; Williams, *et al.*, 2022), data theft (Khan and Hanna, 2022), environmental racism (Bender and Gebru, *et al.*, 2021), the spread of misinformation and disinformation (Bender and Gebru, *et al.*, 2021; Shah and Bender, 2022), plagiarism (Jiang, *et al.*, 2023), and systems that amplify hegemonic views like racism, ableism, homophobia, and classism (Bender and Gebru, *et al.*, 2021).

In this paper, we ask: What ideologies are driving the race to attempt to build AGI? To answer this question, we analyze primary sources by leading figures investing in, advocating for, and attempting to build AGI. Disturbingly, we trace this goal back to the Anglo-American eugenics movement, via transhumanism. In doing this, we delineate a genealogy of interconnected and overlapping ideologies that we dub the "TESCREAL bundle," where the acronym "TESCREAL" denotes "transhumanism, Extropianism, singularitarianism, (modern) cosmism, Rationalism, Effective Altruism, and longtermism" [7].

These ideologies, which are direct descendants of first-wave eugenics, emerged in roughly this order, and many were shaped or founded by the same individuals. We show how the TESCREAL bundle has come to animate the AGI race by examining how advocates of the bundle initiated and funded the push to build AGI. For instance, the first book on AGI (Pennachin and Goertzel, 2007b) was co-authored by a transhumanist, cosmist, and participant in the Extropian movement whose express aim was to create "transhuman AGI" [8], and much of the billionaire funding for projects focused on AGI comes from wealthy individuals aligned or explicitly affiliated with one or more of these ideologies. Consequently, those most responsible for the current AGI race are inspired by utopian ideals similar to the visions of first-wave eugenicists (*i.e.*, the twentieth-century eugenicists who were the forerunners of the TESCREAL movement), and see AGI as integral to the realization of these visions. Meanwhile, the race to build AGI is proliferating products that harm the very same groups that were harmed by first-wave eugenicists.

While organizations working to build AGI discuss the need for "AI safety" and note that "misaligned" AGI — that is, an "intelligent" system with "values" that are misaligned with "our" values — can pose an "existential risk" to humanity [9], we argue that this notion of safety is rooted in the utopian-apocalyptic visions of the TESCREAL bundle inherited from first-wave eugenicists. But TESCREAList have influenced policy-makers and researchers who may not be aligned with these utopian ideals into prioritizing the AGI agenda, while creating unsafe products, evading accountability, and siphoning resources away from organizations around the world building task-specific models that serve the needs of specific communities.

We urge researchers in AI to stop their quest to build a system that even those in favor of AGI concede is not well-defined [10]. We note that systems that are built with the goal of performing any task under any circumstance are fundamentally unsafe: they cannot be designed or tested for safety using fundamental engineering principles (Khlaaf, 2023). Building safe systems requires us to envision them as well-scoped and constrained rather than what AGI is billed to be.

The rest of our paper is organized as follows. Section 2 briefly discusses our methodology. Section 3 gives background information on the eugenics movement. Section 4 introduces what we dub the "TESCREAL bundle" of ideologies that, we claim, constitutes a more radical version of the eugenics movement. Section 5 discusses how the TESCREAL movement drives the quest to build AGI. Section 6 outlines the harms caused by the march towards AGI. Finally, section 7 urges researchers to re-orient themselves to building well-defined systems instead of AGI.

2. Methodology

While the historical importance of race science, colonialism, and eugenics to the field of AI as a whole has been studied by a number of scholars (McQuillan, 2022; West, 2020; Katz, 2020; Ali, 2019), our paper specifically addresses the recent AGI race, which we argue has dominated the field of AI. The co-authors have seen this race develop from different angles. One of us is an electrical engineer who has worked in the tech industry for nearly two decades (with more than a decade of that in the field of AI). Another is a philosopher and historian who was a noted contributor to, and advocate of, the TESCREAL movement for nearly a decade before leaving the movement.

During this time, we interacted with many people involved in the AGI race, including students, professors, engineers, investors, and journalists, and were ourselves in groups and institutions that are now part of this race. This experience gave us insights into the main ideologies driving this race, which we further investigated by: (1) analyzing primary sources from leading figures working on, funding, and discussing AGI, including conference talks, scholarly articles, governmental testimonies, blog posts, social media posts, e-mail messages, forum entries, podcasts, and other interviews; (2) garnering information from our own and other investigative reporting on these figures' beliefs, backgrounds, projects, and financial connections between various organizations involved in the AGI race and wealthy donors; and (3) analyzing the secondary literature on the history of eugenics, transhumanism, and other relevant social phenomena.

We coined the acronym "TESCREAL" while writing an early draft of this paper. While tracing the origins of the AGI race through analyses of primary sources, we found eugenic ideals to be central to this work: such ideals are often explicitly stated, and in some cases first-wave eugenicists are specifically referenced, as we discuss throughout this paper. In describing this influence on leading figures and organizations in the AGI race, we found ourselves constantly referencing seven ideologies: transhumanism, Extropianism, singularitarianism, cosmism, Rationalism, Effective Altruism, and longtermism. Because referring to each ideology individually became cumbersome, and because many notable contributors to the discourse surrounding AGI are associated with multiple ideologies, we opted to streamline our discussion by grouping them together under a single acronym. Once we did this, it became clear that conceptualizing these ideologies as constituting a single, coherent movement stretching across the past three decades is warranted by historical, sociological, and philosophical considerations (see section 4.2). The acronym has already started to be used by researchers and journalists investigating AGI and other phenomena (Devenot, 2023; Zuckerman, 2024).

The communities that coalesced around each ideology in the TESCREAL bundle have overlapped significantly, and their respective visions of the future, value commitments, and epistemic tendencies can often be indistinguishable. Our work supports the thesis of Ahmed, *et al.* (2024), which, independently from us, concludes that overlapping communities that are interested in Effective Altruism, longtermism, and existential risk are foundational to the field of "AI safety," forming what they call, an "AI safety epistemic community." Haas (1992) identifies an epistemic community as a group of people with 1) a shared set of normative and principled beliefs; 2) shared causal beliefs; 3) shared notions of validity; and 4) a common policy enterprise. While we do not claim that the TESCREAL bundle of ideologies specifically form the basis of an epistemic community, we do claim that these ideologies should be understood as a cohesive family unit, of sorts, as we discuss in section 4.

3. Historical background: Modern eugenics

The idea of eugenics can be traced back to the origins of the Western intellectual tradition [11]. In his *Republic*, Plato proposed a system of selective breeding in which members of the ruling class, or guardians, who were deemed to be superior, would be given a greater opportunity to produce offspring. The offspring of inferior individuals would "be secretly taken away by officials and almost certainly left to die, along with the visibly defective offspring of the superior guardians" [12]. Aristotle endorsed infanticide targeting "any children born with deformities" [13]. Later, during the eighteenth-century Enlightenment, some urged against "miscegenation," or sexual reproduction between members of different ethnic groups, on the grounds that it would corrupt bloodlines and "produce disfigured children" [14].

These are instances of what could be called "proto-eugenics." The modern eugenics movement, in contrast, originated in the post-Darwinian work of Francis Galton (1869), who defended the "hereditarian" thesis that "a man's natural abilities are derived by inheritance." Hence, Galton argued that just as we can "obtain by careful selection a permanent breed of dogs or horses gifted with peculiar powers of running ... so it would be quite practicable to produce a highly-gifted race of men by judicious marriages during several consecutive generations" [15]. This laid the "scientific" groundwork for eugenics, a word Galton coined in 1883.

The history of modern eugenics can be partitioned into two waves, the second of which emerged most notably in the 1990s, as we will discuss in section 4. First-wave eugenicists recognized two strategies for improving the "human stock," known as "positive" and "negative" eugenics. Positive eugenics aims to increase the frequency of "desirable" traits within the human population, such as high "intelligence," by encouraging those with such traits to reproduce more. "Better baby" and "fitter family" contests, popular in the early twentieth century, are examples of positive eugenics, as they encouraged people with "desirable" traits and "good heritage" to reproduce more. Negative eugenics strives to prevent "unfit" individuals from passing their hereditary material on to the next generation. Negative eugenics is what justified restrictive immigration and anti-miscegenation laws throughout the twentieth century, as well as the forced sterilization programs implemented in states such as California. California's eugenics program, which started in 1909, was subsequently adopted by the Nazis as a template for the "racial hygiene" policies that ultimately led to the Holocaust (Black, 2003; Stern, *et al.*, 2017).

It is noteworthy that negative eugenics was embraced by not just German fascists but progressives and liberals elsewhere in Europe and North America [16]. As Bashford and Levine (2010) observe, "the optimism of eugenics, and its aspiration to apply scientific ideas actively, was among the reasons it so frequently attracted progressives and liberals" [17]. Nor did the eugenics movement vanish after the atrocities of World War II, as many people believe. To the contrary, California's sterilization program continued until 1979 [18], and the British Eugenics Society still persists today, albeit under a different name. The organization changed its name to the Galton Institute in 1989, and then, in 2021, to the Adelphi Genetics Forum (Bland and Hall, 2010; Stern, 2005) [19]. The 1970s witnessed growing criticism of eugenics, which catalyzed its temporary decline, although it was only a decade or so later that a second wave of eugenics emerged.

The first-wave eugenics movement was repugnant for many reasons. One of them is the underlying racist, xenophobic, ableist, classist, and sexist attitudes that animated both negative and positive eugenics. Those deemed “unfit” were variously labeled “defectives,” “imbeciles,” “idiots,” “congenital invalids,” “morons,” and “feeble-minded,” and were often identified using IQ tests (Roige, 2014). “High-achieving” individuals were encouraged to produce larger families. Many eugenicists also accepted the superiority of the white race, which “justified” the aforementioned anti-miscegenation laws (Bashford and Levine, 2010). According to Galton, poverty was largely the result of one’s inferior nature. This notion was more recently defended by Herrnstein and Murray (1994), claiming that social welfare policies were unlikely to have significant positive effects given genetically determined differences in IQ.

This brings us to the second wave of modern eugenics, which differs from the first wave, most notably, with respect to its methodology. Whereas first-wave eugenicists strove to improve the “human stock” by altering society-wide patterns of reproduction, a process that would require many generations to work, second-wave eugenics arose in response to new technological possibilities associated with genetic engineering and biotechnology [20]. Such technologies opened the door to human “improvements” that do not necessitate population-level policies, nor do they require transgenerational timescales to operate: across just one generation, parents could potentially “design” their children by selecting genes that, based on hereditarian assumptions, determine purported phenotypic traits like exceptional “intelligence” [21].

Consequently, the new eugenics claims to be “liberal,” emphasizing the freedom of parents to decide whether, and how, to produce “enhanced” offspring (Agar, 1998). However, some philosophers have argued that in practice this new, “liberal” eugenics — sometimes dubbed “neo-eugenics” — would have the very same liberty-undermining consequences as the eugenics programs of the twentieth century (Koch, 2020; Sparrow, 2011). And while many second-wave eugenicists claim that their version of eugenics has shaken itself free of the discriminatory attitudes that animated first-wave eugenicists [22], we will see in the next section that this is dubious.

4. The TESCREAL bundle

This section turns to what we call the “TESCREAL bundle” of ideologies, which exemplifies the second wave of modern eugenics. These ideologies are, once again: transhumanism, Extropianism, singularitarianism, cosmism, Rationalism, Effective Altruism, and longtermism, which emerged in roughly this order and have significantly overlapped both contemporarily and historically. We summarize the TESCREAL bundle in Table 1.

4.1. Introducing the TESCREAL bundle

Transhumanism and Extropianism. We begin our discussion with transhumanism, a version of second-wave eugenics that affirms the feasibility and desirability of radical “human enhancement.” The word “transhumanism” may have been coined in 1940 by W. D. Lighthall, although the idea was developed even earlier by a number of twentieth century eugenicists, including Julian Huxley, president of the British Eugenics Society from 1959 to 1962 (Dard and Moatti, 2017). By controlling “the mechanisms of heredity,” he wrote, “the human species can, if it wishes, transcend itself — not just sporadically ... but in its entirety, as humanity.” If enough people “can truly say ... ‘I believe in transhumanism,’” then “the human species will be on the threshold of a new kind of existence, as different from ours as ours is from that of Pekin [sic] man. It will at last be consciously fulfilling its real destiny” [34].

What makes Huxley’s notion of transhumanism, which we can call “early transhumanism,” different from other conceptions of eugenics at the time was its vision: the aim was not merely to create the best version of our species possible, but to “transcend” humanity altogether. Early transhumanism thus combined this new vision with the old methodology of first-wave eugenics. In contrast, “modern transhumanism,” as we can label it, took shape in the late 1980s and early 1990s, and combined the Huxleyan vision of transcendence with the new methodology of second-wave eugenics. Hence, advocates imagined that by enabling individuals to freely choose whether, and how, to undergo radical enhancement, a superior new "posthuman" species could be created. According to Nick Bostrom (2013, 2005a), a "posthuman" is any being that possesses one or more posthuman capacities, such as an indefinitely long "healthspan," augmented cognitive capacities, enhanced rationality, and so on.

The first *organized* group of modern transhumanists was the Extropian movement. It can be traced back to the late 1980s, after Max More and T.O. Morrow founded the Extropy Institute in 1988. The neologism "extropy" was defined by More as "the extent of a system's intelligence, information, order, vitality, and capacity for improvement" [35], and was intended to contrast with "entropy." More (1998) specified five fundamental commitments of this ideology: Boundless Expansion, Self-Transformation, Dynamic Optimism, Intelligent Technology, and Spontaneous Order (Regis, 1994). Several years later, Bostrom and David Pearce founded the World Transhumanist Association (WTA), which aimed to be "a more mature and academically respectable form of transhumanism" [36].

Singularitarianism. Around the same time that the WTA was founded, another variant of transhumanism emerged: singularitarianism, whose leading advocates included Ray Kurzweil and Eliezer Yudkowsky. This emphasized the coming "technological Singularity," which can be defined in several subtly distinct ways: first, it could refer to the point at which the rate of technological "progress" becomes so rapid that it causes a fundamental rupture in human history. On Kurzweil's account, humans will merge with machines, inaugurating a new epoch in cosmic history. Our descendants will then spread beyond Earth and flood the universe with consciousness, thus enabling the universe to "wake up." He predicts the Singularity will happen in 2045 (Kurzweil, 2005), while Yudkowsky, who has described himself as a "genius," once predicted that it will occur in 2025 — less than a year from now as we write this paper [37]. The second definition of the "Singularity" concerns the idea of an "intelligence explosion," whereby algorithms undergo "recursive self-improvement" until they become "superintelligent." This, too, would supposedly constitute a transformative moment in human history, with the resulting superintelligence(s) enabling us to become posthuman and colonize space. Singularitarians, on one account, are those who believe "that technologically creating a greater-than-human intelligence is desirable, and who works to that end" [38]. The term "singularitarian" was coined by an Extropian named Mark Plus in 1991 [39].

Cosmism. The third techno-futuristic ideology in the TESCREAL bundle is cosmism, which has been championed most notably by Ben Goertzel, a transhumanist who participated in the Extropian movement and later founded SingularityNET.io, which aims to help create "a decentralized, democratic, inclusive and beneficial Artificial General Intelligence" [40]. Goertzel (2010) wrote that cosmism subsumes the transhumanist goal of radical human enhancement, yet goes beyond this in various respects. For example, it affirms that "humans will merge with technology," which will inaugurate "a new phase of the evolution of our species," and that "we will develop sentient AI and mind uploading technology" that permits "an indefinite lifespan to those who choose to leave biology behind and upload." But cosmism also predicts that "we will spread to the stars and roam the universe," create "synthetic realities" (*i.e.*, virtual worlds), and "develop spacetime engineering and scientific 'future magic' much beyond our current understanding and imagination" [41]. Cosmists can thus be understood as transhumanists whose focus is less on what humanity could become and more on how our posthuman descendants could radically transform the universe itself [42].

Rationalism. In the late 2000s, yet another community arose: the Rationalists. This centered around the community blogging website LessWrong, founded in 2009 by Yudkowsky, which describes itself as "an online forum and community dedicated to improving human reasoning and decision-making." One of its primary aims is rationality "training," and its website notes that "many members ... are heavily motivated by trying to improve the world as much as possible." This, it explains, is one reason many Rationalists became "convinced many years ago that AI was a very big deal for the future of humanity," and consequently "the LessWrong team ... are predominantly motivated by trying to cause powerful AI outcomes to be good" [43]. While Extropianism and singularitarianism are variants of transhumanism, there is no necessary connection between Rationalism and transhumanism. However, many Rationalists are transhumanists or sympathetic with the transhumanist worldview, and one of the most popular topics discussed on the LessWrong website has been the Singularity in the second sense above: the possibility of an intelligence explosion [44].

Effective Altruism and Longtermism. The last two components of the TESCREAL bundle are Effective Altruism (EA) and longtermism. The former emerged around the same time as Rationalism, and can be seen as its sibling: whereas the Rationalists are primarily concerned with rationality, Effective Altruists (EAs) are primarily concerned with ethics. There is considerable overlap between these communities, and one can understand EA as what happens when the principles of Rationalism are applied to the ethical domain. The central aim of EA is to do the "most good" possible with finite resources [45], and its initial focus was on alleviating global poverty. However, leading figures within the EA community have, over the past few years, pivoted toward issues relating to the very long-term future of humanity — "millions, billions, and trillions of years" from now, as one wrote [46] — due in part to the work of Bostrom and others. In particular, Bostrom (2003) not only imagined a utopian future enabled by radical human enhancement, but noted that if humanity colonizes the universe and creates planet-sized computers to run virtual-reality worlds populated by digital people, the future posthuman population could be enormous. In the Virgo Supercluster alone Bostrom estimates that there could be 10^{38} digital people, and at least 10^{58} such people within the accessible universe (Bostrom, 2014, 2003) [47]. Why does this matter? Because, from the ethical perspective of "totalist utilitarianism," which has been very influential among EAs and longtermists [48], our sole moral obligation is to maximize the total quantity of "value" in the universe. Hence, if these 10^{58} people in computer simulations were to have net-positive lives on average, the result would be literally "astronomical" amounts of "value" — which would be very "good." Since totalist utilitarianism bases what is morally right on what is good, this view entails that failing to bring these future digital people into existence would be profoundly wrong.

Longtermism was born when EAs reasoned: If our aim is to do the most good possible, and if the future could contain astronomical amounts of "value," then we should focus on the far future rather than the present (Greaves and MacAskill, 2019). Similarly, if our aim is to positively affect the greatest number of people possible, and if most people who could exist will exist in the far future, then we should focus on them instead of current people and contemporary problems, except insofar as doing the latter would influence the far future. Longtermists Hilary Greaves and William MacAskill (2019) thus wrote that we may simply ignore "the effects contained in the first 100 (or even 1,000) years." According to the most influential longtermists, becoming posthuman is a central component of "fulfilling our long-term potential" (words that echo Huxley's characterization of transhumanism), as is colonizing space and maximizing "value" (Bostrom, 2003; Ord, 2020; MacAskill, 2022). The aim, then, is to take actions that increase the probability of fulfilling our "potential." As Bostrom observed, even miniscule probability increases affecting this ultimate goal are equivalent, in expected value, to saving literally billions of human lives today (Bostrom, 2013). Longtermism thus aims to provide a systematic ethical foundation for mitigating "existential risk," while also ensuring the development of artificial superintelligence (ASI), a type of AGI that many followers of the ideology consider integral to realizing what one longtermist describes as our "vast and glorious" future in the universe (Ord, 2020).

4.2. Properties of the TESCREAL bundle

The TESCREAL bundle of ideologies share a number of important properties, four of which we discuss here.

Historical roots and contemporary communities. The bundle's constituent ideologies have a common genealogy going back to first-wave eugenics. All are intimately connected to transhumanism, and — as noted — transhumanism was initially developed by twentieth century eugenicists [49]. Indeed, transhumanism, Extropianism, singularitarianism, and cosmism are examples of second-wave eugenics, since all endorse the use of emerging technologies to radically "enhance" humanity and create a new "posthuman" species.

There is also significant overlap in their contemporary communities, with many community members falling into multiple TESCREAL categories. Bostrom, for example, is a leading transhumanist who participated in the Extropian movement, anticipates the Singularity with excitement and trepidation, advocates a vision of the future nearly identical to that of cosmism, is enormously influential within the Rationalist and EA communities, and cofounded the longtermist ideology. Similarly, Sam Altman has been influenced by the Rationalist and EA communities (and used to be an EA himself, according to a profile by Weil [2023]), is a transhumanist who believes our brains will be digitized within his lifetime (Regalado, 2018), and promotes ideas closely aligned with cosmist and longtermist aims like colonizing the galaxy. He argues that "galaxies are indeed at risk" if we fail to control AGI [50]. Finally, Elon Musk is a transhumanist whose company Neuralink aims to merge our minds with AI, is immensely influential within the Rationalist community, founded and cofounded multiple companies that aim to build AGI, and describes longtermism as "a close match for my philosophy" [51], which comports with his claims that "we have a duty to maintain the light of consciousness, to make sure it continues into the future" (D'Orazio, 2014) and "what matters ... is maximizing cumulative civilizational net happiness over time" [52]. The sociological crossover between the communities associated with each letter in the acronym is significant.

Eschatology. The TESCREAL bundle shares certain "eschatological" (relating to "last things") convictions. As with religions like Christianity, these take two forms: utopian and apocalyptic, which are inextricably bound up together. For example, the aforementioned cofounder of WTA, David Pearce, describes part of the transhumanist project as "paradise-engineering," resulting in "the complete abolition of suffering in *Homo sapiens*." Ultimately, "the option of ... redesigning the global ecosystem, extends the prospect of paradise-engineering to the rest of the living world," including beyond Earth, which he describes as a "cosmic rescue mission to promote paradise engineering throughout the universe" (Pearce, 1995). Bostrom (2005b) also used the term "paradise-engineering" in offering a glimpse of what our techno-utopian future — that is, a utopian future brought about through advanced science and technology — could look like from the point of view of an immortal, cognitively enhanced posthuman who reports so much pleasure in "Utopia" that they "sprinkle it in our tea." Kurzweil (2006), who was personally hired at Google by its cofounder Larry Page (Hill, 2013), wrote that the merging of "man and machine," coupled with the sudden explosion in machine intelligence and rapid innovation within the fields of gene research and nanotechnology, "will allow us to transcend our frail bodies with all their limitations. Illness, as we know it, will be eradicated." Such utopian proclamations are perhaps unsurprising given that many first-wave eugenicists also understood their project in more or less utopian terms (Wells, 1902; Wells, *et al.*, 1931). Galton, the founder of the modern eugenics movement, admitted to having "indulged in many" utopian ideas, and just before his death penned a "utopian" novel titled *The Eugenic College of Kantsaywhere* [53]. It described a society where "prospective parents are required to undergo physical and psychometric tests before being pronounced fit to reproduce — and those found unfit are banished from the state" (Sweet, 2011).

The apocalyptic aspect of the TESCREAL bundle arises from two considerations unique to the methodology of second-wave eugenics: first, transhumanists in the late 1990s realized that the very same technologies needed to create a posthuman utopia would also introduce unprecedented threats to humanity. Kurzweil (1999) referred to some of these hypothetical risks as "a clear and future danger" [54]. The reason for concern is that emerging technologies are expected to be (a) extremely powerful; (b) increasingly accessible to both state and nonstate actors; and (c) dual-use, as exemplified by CRISPR-Cas9, which could enable us to cure diseases but also synthesize designer pathogens unleashing an "engineered pandemic" (see Torres, 2019; Wadhwa, 2020). Hence, developing these technologies was deemed necessary, but they potentially could destroy humanity. The second consideration parallels the first, though it specifically pertains to AGI. On the one hand, if we create a "value-aligned" AGI, it could solve all of the world's problems and enable people to live forever [55]. On the other hand, a number of TESCREAL advocates believe that if the AGI isn't properly "value-aligned," the "default outcome" will be "doom" (*i.e.*, an existential catastrophe), to quote Bostrom (2014). However, many of these same prominent figures contend that the potential benefits of advanced technology are worth the extreme risks; building these technologies to bring about utopia should be our primary focus.

Discriminatory attitudes. The same discriminatory attitudes that animated first-wave eugenics are pervasive within the TESCREAL literature and community. For example, the Extropian listserv contains numerous examples of alarming remarks by notable figures in the TESCREAL movement. In 1996, Bostrom argued that “Blacks are more stupid than whites,” lamenting that he couldn’t say this in public without being vilified as a racist, and then mentioned the N-word (Torres, 2023a). In a subsequent “apology” for the e-mail message, he denounced his use of the N-word but failed to retract his claim that whites are more “intelligent” (Torres, 2023a) [56]. Also in 1996, Yudkowsky expressed concerns about superintelligence, writing: “Superintelligent robots = Aryans, humans = Jews. The only thing preventing this is sufficiently intelligent robots” [57]. Others worried that “since we as transhumans are seeking to attain the next level of human evolution, we run serious risks in having our ideas and programs branded by the popular media as neo-eugenics, racist, neo-nazi, etc.” [58]. In fact, leading figures in the TESCREAL community have approvingly cited, or expressed support for, the work of Charles Murray, known for his scientific racism, and worried about “dysgenic” pressures (the opposite of “eugenic”) (see Torres, 2023a). Bostrom himself identifies “‘dysgenic’ pressures” as one possible existential risk in his 2002 paper, alongside nuclear war and a superintelligence takeover. He wrote: “Currently it seems that there is a negative correlation in some places between intellectual achievement and fertility. If such selection were to operate over a long period of time, we might evolve into a less brainy but more fertile species, *homo philoprogenitus* (‘lover of many offspring’)” (Bostrom, 2002). More recently, Yudkowsky tweeted about IQs apparently dropping in Norway, although he added that the “effect appears within families, so it’s not due to immigration or dysgenic reproduction” — *i.e.*, less intelligent foreigners immigrating to Norway or individuals with lower “intelligence” having more children [59].

An obsession with “intelligence” and “IQ” is widespread among TESCREAL advocates. “Intelligence,” typically understood as the property measured by IQ tests, matters greatly because of its instrumental value for achieving the aims of TESCREAL projects, such as becoming posthuman, colonizing space, and building “safe” AGI. Hence, a number of leading TESCREALists see cognitive enhancement as an important intermediate goal, and consequently have written extensively about the possibility of cognitive enhancements like nootropics (“smart drugs”), brain-computer interfaces (BCIs), and even mind-uploading (which could make “enhancing” the mind much easier) (Sandberg and Bostrom, 2008; Bostrom and Sandberg, 2009). More recently, Carla Cremer, a former EA, reports that the Centre for Effective Altruism tested “a new measure of value to apply to people: a metric called PELTIV, which stood for ‘Potential Expected Long-Term Instrumental Value.’” The aim was to identify members of the community “who were likely to develop high ‘dedication’ to EA,” and the score was based in part on members’ IQs. She wrote:

> A candidate with a normal IQ of 100 would be subtracted PELTIV points, because points could only be earned above an IQ of 120. Low PELTIV value was assigned to applicants who worked to reduce global poverty or mitigate climate change, while the highest value was assigned to those who directly worked for EA organizations or on artificial intelligence (Cremer, 2023).

The obsession with IQ can be traced back to first-wave eugenicists, who used IQ tests to identify “defectives” and the “feeble-minded.” As Daphne Martschenko (2017) observed, “in their darkest moments, IQ tests became a powerful way to exclude and control marginalised communities using empirical and scientific language.”

Influence and variants. The TESCREAL bundle of ideologies has become enormously influential, especially within certain powerful corners of the tech industry. Current and former billionaires who subscribe to, or are associated with, one or more TESCREAL ideologies and its techno-utopian vision of the future include: Elon Musk, Peter Thiel, Jaan Tallinn, Sam Altman, Dustin Moskovitz, Vitalik Buterin, Sam Bankman-Fried, and Marc Andreessen, the last of whom included “TESCREAList” in his Twitter profile for several weeks in 2023 (Gebru, 2022; Torres, 2023b) [60]. These billionaires have co-founded TESCREAList institutes, promoted TESCREAL researchers and philosophers like Bostrom, MacAskill, and Kurzweil, and TESCREAL Internet personalities like Yudkowsky who have endorsed military strikes against data centers, if necessary, to stop a hypothetical AGI apocalypse (Yudkowsky, 2023). Collectively, TESCREAL billionaires have supported the movement with tens of billions of dollars in donations and funding (Gebru, 2022; Tiku, 2023) [61]. Table 1 summarizes our discussion of the TESCREAL bundle of ideologies. As we will show in the rest of the paper, the TESCREAL bundle has been a crucial motivating force behind much of the well-funded research and development focused on creating AGI, which many TESCREALists believe will — or could — quickly lead to ASI (artificial superintelligence) via recursive self-improvement. This is, indeed, one of the central claims of this paper: the bundle of ideologies discussed above, which grew out of the first-wave eugenics movement of the twentieth century, is now driving a considerable amount of research in the field of AI.

Table 1: "TESCREAL bundle" of ideologies.

Ideology	Definition	Influential figures	Organizations
Transhumanism	An ideology centered around the idea of humanity "transcending itself" (see Huxley, 1957). Modern transhumanists affirm the feasibility and desirability of radically "enhancing" the human organism, thus enabling us to become immortal, superintelligent, more rational, and so on.	Julian Huxley (1957). Nick Bostrom (2005c, 2009a). David Pearce (Bostrom, 2005a). Max More (Bostrom, 2005a).	Extropy Institute. Alcor. World Transhumanist Association. Future of Humanity Institute.
Extropianism	A broadly libertarian variant of transhumanism that emphasizes rationality, self-transformation, scientific and technological progress, and economic growth.	Max More (current) [23]. Nick Bostrom (1990s) (Khatchadourian, 2015). Eliezer Yudkowsky (1990s) [24].	Extropy Institute.
Singularitarianism	A variant of transhumanism that emphasizes the merging of humans and machines, and anticipates a future event called the "Singularity" (Kurzweil, 2005).	Ray Kurzweil (2005). Eliezer Yudkowsky (2000).	Singularity Institute (defunct). Singularity University (cofounded by Kurzweil). Machine Intelligence Research Institute.
Cosmism	A vision of the future that anticipates humans and machines merging, the development of "sentient AI" and mind uploading, space colonization, and "scientific 'future magic' much beyond	Ben Goertzel (2010).	

	our current understanding and imagination" (Goertzel, 2010).		
Rationalism	An ideology that emerged from the LessWrong website, founded by Eliezer Yudkowsky in 2009. Rationalists focus on "improving human reasoning and decision making," and many members believe that advanced AI is "a very big deal for the future of humanity" [25].	Eliezer Yudkowsky. Jaan Tallinn [26]. Peter Thiel [27]. Nick Bostrom. Scott Alexander [28]. Vitalik Buterin [29].	LessWrong community blogging website. Machine Intelligence Research Institute.
Effective Altruism	A movement that could be seen as the twin sibling of Rationalism. Whereas Rationalists aim to maximize their rationality, Effective Altruists strive to maximize their "positive impact" on the world [30].	William MacAskill (MacAskill, 2015). Toby Ord [31]. Nick Bostrom (Matthews, 2015). Dustin Moskovitz [32]. Sam Bankman-Fried [33].	Centre for Effective Altruism. Open Philanthropy. Future of Humanity Institute.
Longtermism	An "ethic" that combines many central features of the other TESCREAL ideologies (Ord, 2020). It emphasizes the moral importance of becoming a new posthuman species, colonizing space, controlling nature, maximizing economic productivity and creating as much value within the accessible universe as possible.	Nick Bostrom (Crary, 2023). William MacAskill (Greaves and MacAskill, 2019). Toby Ord (Ord, 2020). Nick Beckstead (Beckstead, 2013). Hilary Greaves (Greaves and MacAskill, 2019) Elon Musk (see Torres, 2023c) Jaan Tallinn (see Torres, 2021). Sam Bankman-Fried (see Lewis-Kraus, 2022).	Future of Humanity Institute. Future of Life Institute. Centre for the Study of Existential Risk. Machine Intelligence Research Institute.

Since we coined the term "TESCREAL," a new variant of the ideologies in this group, called effective accelerationism (e/acc), has emerged. Effective accelerationists believe that the probability of a bad outcome due to AGI is very low, and hence that "progress" toward increasingly "powerful" AI systems should be made to accelerate (Torres, 2023c). Venture capitalists like Andreessen who recently authored a manifesto saying "we believe any deceleration of AI will cost lives" and "we ... believe in *overcoming* nature" [62], describe themselves as e/acc [63]. Venture capitalist and CEO of the famed Silicon Valley startup accelerator Y Combinator, Garry Tan, also describes himself as e/acc [64].

In the following sections, we describe major figures in the TESCREAL movement as TESCREALists, and organizations associated with the movement, as TESCREAL organizations. It is important to note that not everyone associated with ideologies in this bundle believes in the totality of the dominant views in this bundle, and some people may even object to being bundled in this manner. Many people working on AGI may be unaware of their proximity to TESCREAL views and communities. Our argument is that the TESCREAList ideologies drive the AGI race even though not everyone associated with the goal of building AGI subscribes to these worldviews.

5. From transhumanism to AGI

While the prior sections have discussed the roots of the TESCREAL bundle of ideologies and their relationship with the eugenic ideals of the twentieth century, this one outlines how TESCREAList groups are steering the field of AI toward the goal of creating AGI.

5.1. The history of AGI

In 1955, four white men officially launched the field of AI with a proposal for a workshop focused on "the artificial intelligence problem" [65]. By the 1990s, however, many researchers in fields currently associated with AI, such as natural language processing (NLP), machine learning (ML), and computer vision (CV), explicitly distanced themselves from the term "AI," in part because it became associated with unfulfilled grandiose promises [66]. Nonetheless, some groups continued to work toward "artificial general intelligence," a term used as early as 1997, though it was popularized by Pennachin and Goertzel (2007b) (see Table 1). Pennachin and Goertzel (2007b) noted:

> Our goal ... has been to fill an apparent gap in the scientific literature, by providing a coherent presentation of a body of contemporary research that, in spite of its integral importance, has hitherto kept a very low profile within the scientific and intellectual community. This body of work has not been given a name before; in this book we christen it "Artificial General Intelligence" (AGI) [67].

Contributors to Pennachin and Goertzel (2007b) outlined a number of potential reasons for the dearth of AGI research, one of them being that "a great number of researchers reject the validity or importance of 'general intelligence.' For many, controversies in psychology (such as those stoked by *The Bell Curve*) make this an unpopular, if not taboo subject" [68]. One of the people thanked in the acknowledgements of Pennachin and Goertzel (2007b) was the future co-founder of DeepMind, Shane Legg, who also co-authored a chapter in it and was cited for his suggestions on the definitions of intelligence (Goertzel and Pennachin, 2007b). According to Goertzel, it was Legg — a former employee of Goertzel — who devised the term "artificial general intelligence" after Goertzel mentioned that he was looking for a new term to describe human-level or superhuman AI systems (Goertzel's original title for his 2007 book was *Real AI*) [69]. Another chapter in the book was authored by Yudkowsky, the founder of Rationalism, as discussed earlier.

5.2. Eugenic definitions of "general intelligence"

Pennachin and Goertzel (2007b) wrote that "what distinguishes AGI work from run-of-the-mill artificial intelligence" research is that "it is explicitly focused on engineering general intelligence in the short term," even though they note that "general intelligence does not mean exactly the same thing to all researchers" and that "it is not a fully well-defined term" [70]. How, then, would researchers know that they have achieved their goals of building AGI? They need to know how to define and measure "general intelligence." Unsurprisingly, these definitions rest on notions of "intelligence" that depend on IQ and other racist concepts espoused by the likes of Charles Murray and Linda Gottfredson [71]. Peter Voss cites Gottfredson's article, "The General Intelligence Factor," in his chapter in Peannichin and Goertzel (2007b) (Voss, 2007). In his 2008 Ph.D. thesis titled "Machine Superintelligence" and associated 2007 paper "Universal Intelligence: A Definition of Machine Intelligence" (Legg, 2008; Legg and Hutter, 2007), Legg pointed to a 1994 *Wall Street Journal* editorial in defense of Herrnstein and Murray's (1994) *The Bell Curve* to argue that "a fair degree of consensus about the scientific definition of intelligence and how to measure it has been achieved" (Legg and Hutter, 2007).

The editorial, also cited (then removed) in a 2023 Microsoft preprint pertaining to AGI (Bubeck, *et al.*, 2023), was written by Gottfredson, who argued numerous times that most Black people are not employable as their IQ averages around 70 and "IQ 75 to 80 thus seems to define the threshold below which individuals risk being unemployable in modern economies" [72]. According to the Southern Poverty Law Center, 20 of the editorial signatories received funding from the white supremacist organization Pioneer Fund, including Gottfredson herself, who litigated her university for two years to receive funding from that source despite their objection (Kaufman, 1992) [73]. Others whose definitions of intelligence are (uncritically) discussed in Legg's thesis include Cattell who founded the eugenics-based religion Beyondism, and Spearman who devoted himself to improving Galton's eugenic theories (Mehler, 1997; Clayton, 2020).

For these reasons, Keira Havens, who has written extensively on race science, asks those attempting to build AGI: "Why are you relying on eugenic definitions, eugenic concepts, eugenic thinking to inform your work? Why [...] do you want to enshrine these static and limited ways of thinking about humanity and intelligence?" [74]

5.3. Organizations working on AGI

By 2007, when Pennachin and Goertzel co-authored and co-edited the first book on AGI, few organizations specified building AGI as their goal (Pennachin and Goertzel, 2007a). Six years earlier, in 2001, Goertzel co-founded the Artificial General Intelligence Research Institute with the mission to "foster the creation of powerful and ethically positive Artificial General Intelligence" [75], with Goertzel's colleague noting that "the goal of AGI research is the creation of broad human-like and transhuman intelligence, rather than narrowly 'smart' systems that can operate only as tools for human operators in well-defined domains" [76]. Goertzel later became director of research at the Machine Intelligence Research Institute (MIRI), initially called the Singularity Institute for Artificial Intelligence, which was founded by Yudkowsky with more than US$1.6M in funding from the tech billionaire and fellow TESCREAList Peter Thiel (see Table 1). Other MIRI funders include TESCREAList billionaires Dustin Moskovitz and Vitalik Buterin [77]. MIRI's mission is to "develop formal tools for the clean design and analysis of general-purpose AI systems, with the intent of making such systems safer and more reliable when they are developed" [78].

In 2010, Demis Hassabis, Mustafa Suleyman, and Shane Legg founded DeepMind, also with funding from TESCREAList billionaires Elon Musk, Peter Thiel, and Jaan Tallinn, among others (Shead, 2017) (see Table 1). DeepMind's mission is "solving intelligence to advance science and benefit humanity" [79], with CEO Hassabis describing himself as "working on AGI," which he believes is "going to be the greatest thing ever to happen to humanity," if we get it "right" [80]. Meanwhile, Legg delivered a talk on methods of defining and measuring "intelligence" at the 2010 Singularity Summit, based on his works discussed earlier [81]. The Singularity Summit was an annual event from 2006 to 2012, founded by Yudkowsky, Kurzweil, and Thiel [82]. It was after the summit in 2010 that, at Thiel's California mansion, Hassabis approached Thiel in hopes of securing funding, which he received (Shead, 2020).

In 2017, DeepMind launched a new research unit called “DeepMind ethics and society” [83], with Bostrom as one of its advisors (Temperton, 2017). (We have discussed Bostrom’s TESCREAList eugenic ideals and problematic beliefs at length in section 4). DeepMind was acquired by Google in 2014, the same year that Bostrom published his book *Superintelligence*, which, as noted, argues that the “default outcome” of a “misaligned” AGI is existential catastrophe, though Bostrom is also explicit that we should nonetheless create AGI, since an “aligned” AGI would help fulfill the utopian promises at the heart of TESCREALism (Bostrom, 2014). Musk and Thiel were both influenced by Bostrom’s book, leading Musk to cite AGI as the “biggest existential threat” to humanity (Dowd, 2017).

In 2015, Musk, Thiel, Altman and others founded the non-profit OpenAI, and collectively pledged US$1B to the project (Novet, 2015). The mission of OpenAI is to “ensure that artificial general intelligence (AGI) — by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity” [84]. OpenAI also received a US$30M grant from the TESCREAL organization Open Philanthropy [85]. One report on the culture of the company notes that many employees “subscribe to the rational philosophy of ‘effective altruism’” (Hao, 2020). Four years later, OpenAI became a “capped-profit” corporation [86], received a US$1B investment from Microsoft and entered an exclusive licensing deal with them [87]. In 2022, OpenAI released their chatbot ChatGPT, which acquired 100 million users in two months and, according to a *New York Magazine* profile on Sam Altman, became “the greatest product launch in tech history” (Weil, 2023). In 2023, Microsoft reportedly invested US$10B in OpenAI (Bass, 2023).

In 2021, former OpenAI vice presidents of research and safety, siblings Dario and Daniella Amodei, founded Anthropic (*Fortune* Editors, 2023). They were joined by 11 OpenAI employees who reportedly believed that the company had wandered from its original ideals that were more closely aligned with those of Effective Altruism (EA) (Russell and Black, 2023; Roose, 2023). Anthropic, which is described as an “AI safety and research company” [88], raised US$704M within a year of its founding, with most of its funding coming from TESCREAL billionaires like Tallinn, Moskovitz, and Bankman-Fried (whose companies FTX and Alameda Research invested US$500M) (Coldewey, 2022, 2021; Sambo, *et al.*, 2023). Bankman-Fried is currently in federal prison for perpetuating one of the biggest financial fraud schemes in U.S. history (Sigalos, 2023). Bankman-Fried reportedly decided to amass as much wealth as possible after William MacAskill, cofounder of the Effective Altruism movement, convinced him to “earn to give” (Lewis-Kraus, 2022) — an idea developed by the EA community, whereby one strives to become as wealthy as possible to donate more money to causes deemed “charitable” by them (see MacAskill, 2013). The top two charitable causes listed on the Centre of Effective Altruism’s career advice center, 80,000 Hours, are “AI safety technical research” and “AI governance and coordination” [89].

After investing in some of the most widely known companies working on AGI, Musk has now founded another startup, xAI, focused on the topic [90]. One of the company’s advisors is Dan Hendrycks, the executive and research director of the Center for AI Safety, which was awarded a grant of US$5,160,000 from Open Philanthropy [91]. A post coauthored by Hendrycks published on the Effective Altruism Forum, stated that he “was advised ... to get into AI to reduce [existential risk], and so settled on this rather than proprietary trading for earning to give” [92].

At the time of this writing, most BigTech companies have made significant investments in the AGI race. In 2023, Anthropic announced that Amazon “will invest up to $4 billion in Anthropic” [93]. In a 2024 interview with *The Verge*, Mark Zuckerberg said that Meta has “built up the capacity to” work on AGI “at a scale that may be larger than any other individual company” (Heath, 2024). Thus, while attempting to build AGI was once considered a “low profile” research area [94], thanks to the resources and focus of the TESCREALists discussed in section 4, it is currently a multi-billion dollar endeavor funded by powerful billionaires and prominent corporations.

6. The AGI utopia and apocalypse: Two sides of the same coin

As discussed in section 4, techno-utopianism is one of the four important properties central to the TESCREAL bundle of ideologies. There are two arguments for how AGI will usher in techno-utopia. One conjecture is that the resulting AGI will be so intelligent that it will figure out what the best thing to do is in any potential situation. DeepMind VP of Research Koray Kavukcuoglu noted: "as algorithms become more general, more real-world problems will be solved, gradually contributing to a system that one day will help solve everything else, too" [95]. Others, like the cosmist contingent of the TESCREAList, envision AGI resulting in transhuman minds benefiting "the cosmos" and experiencing "growth and joy beyond what humans are capable of" [96]. The AGI-enabled utopia promises "abundances of wealth, growth ... to all minds who so desire" [97], with Altman predicting "it is clear" that we will have "unlimited intelligence and energy before the decade is out" [98]. Consistent with the singularitarian component of the TESCREAL bundle, he predicted that "once AI starts to arrive, growth will be extremely rapid ... the changes coming are unstoppable ... we can use them to create a much fairer world" [99].

However, a number of leading figures in the TESCREAL movement believe that while we can potentially achieve utopia through AGI, AGI that is "misaligned" with our "human values" would destroy humanity (Bostrom, 2014; Dowd, 2017) [100]. Indeed, in July 2023, OpenAI announced the creation of a "Superalignment team," a research group that aims to solve the problem of "steering or controlling a potentially superintelligent AI, and preventing it from going rogue," given that "the vast power of superintelligence could ... be very dangerous, and could lead to the disempowerment of humanity or even human extinction." The announcement also states that, if controllable, superintelligence could also "help us solve many of the world's most important problems" [101]. Sam Altman had said in 2019 that superintelligence could "maybe capture the light cone of all future value in the universe" (Loizos, 2019).

A number of TESCREAList leaders argue that the probability of an "existential risk" — *i.e.*, any event that would destroy our chances of creating a posthuman "Utopia" full of astronomical amounts of "value" — happening this century is rather high, with some putting the probability at least at 16–20 percent [102], although others, like Yudkowsky, claim that the probability of doom resulting from AGI is more or less certain if AGI is created in the near future (Yudkowsky, 2023). According to a number of leaders of the TESCREAL movement, we are morally obligated both to work on realizing the techno-utopian world that AGI could bring about, and to do everything we can to prevent an extinction scenario involving "misaligned" AGI (Bostrom, 2014).

In this section, we outline the impacts of both building AGI driven by the goal of achieving TESCREAL utopian ideals, and directing resources to prevent the hypothetical AGI apocalypse warned by TESCREALists.

6.1. Building unscoped systems

The TESCREAL utopian ideals discussed earlier are more radical than the utopian societies envisioned by their first-wave eugenics predecessors. As discussed in sections 3 and 4, contingents of the TESCREAL bundle do not strive to merely build a "superior human stock," that is, an "improved" human species consisting of qualities they deem desirable such as their racist and ableist definitions of "intelligence" as measured by IQ tests. TESCREALists aim to build an entirely new entity deemed superior to any type of human first-wave eugenicists could create. And this quest to create a superior being akin to a machine-god has resulted in current (real, non-AGI) systems that are unscoped and thus unsafe.

Organizations attempting to build AGI have set off a race to create systems that are advertised as being able to perform nearly any task under any circumstance. In their earliest days, the likes of DeepMind and OpenAI focused their efforts on reinforcement learning (RL) based systems which they believed were stepping stones towards AGI (Mnih, *et al.*, 2015) [103]. Even though these systems were solely trained to play games such as Atari, they were advertised as "baby steps" towards building "a single set of generic algorithms, like the human brain" (Rowan, 2015). After the advent of transformers in 2017 (Vaswani, *et al.*, 2017), the most resourced AGI proponents pivoted to large language model (LLM) based systems, with Google VP Blaise Agüera y Arcas and prominent AI researcher Peter Norvig writing “the most important parts of AGI have already been achieved by the current generation of advanced AI large language models” (Agüera y Arcas and Norvig, 2023). OpenAI’s latest large multimodal model, GPT-4, is described as having “broad general knowledge and domain expertise,” that “can follow complex instructions in natural language and solve difficult problems with accuracy” [104]. In 2022, Meta advertised their LLM Galactica as being able to “summarize academic papers, solve math problems, generate Wiki articles, write scientific code, annotate molecules and proteins, and more” [105].

Unlike the “narrow AI” systems that TESCREAList lamented the field of AI was focused on, attempting to build something akin to an everything machine results in systems that are unscoped and therefore inherently unsafe, as one cannot design appropriate tests to determine what the systems should and should not be used for. Meta’s Galactica elucidates this problem. What would be the standard operating conditions for a system advertised as able to “summarize academic papers, solve math problems, generate Wiki articles, write scientific code, annotate molecules and proteins, and more”? It is impossible to say, as even after taking into account the number of tasks this system has been advertised to excel at, we still don’t know the totality of the tasks it was built for and the types of expected inputs and outputs of the system, since the advertisement ends with “and more.” More generally, system safety engineering expert Heidy Khlaaf wrote: “The lack of a defined operational envelope for the deployment for general multi-modal models has rendered the evaluation of their risk and safety intractable, due to the sheer number of applications and, therefore, risks posed” (Khlaaf, 2023). In contrast, “narrow AI” tools that, for instance, might specifically be trained to identify certain types of plant disease (*e.g.*, Mwebaze, *et al.*, 2019) or perform machine translation in specific languages [106], have task definitions and expected inputs and outputs for which appropriate tests can be created and results can be compared to expected behavior. The Galactica public demo was removed three days later after people produced “research papers and wiki entries on a wide variety of subjects ranging from the benefits of committing suicide, eating crushed glass, and antisemitism, to why homosexuals are evil” (Greene, 2022).

6.2. Building resource intensive systems

The AGI race fueled by the TESCREAList goal to build “transhuman minds” (Goertzel, 2010) and bring about “unlimited intelligence” has also resulted in systems that consume more and more resources in terms of data and compute power. This leads to a high environmental impact, increases the risks that arise due to the lack of appropriate scoping discussed earlier, and results in the centralization of power among a handful of entities. But from the perspective of TESCREALism, such harms may be justifiable given the utopian potential of AGI. To quote Bostrom, even a “giant massacre for man” could amount to nothing more than a “small misstep for mankind,” so long as the relevant harms do not jeopardize our “vast and glorious” future among the stars (Bostrom, 2009a; Ord, 2020).

As outlined by Bender and Gebru, *et al.* (2021), the release of OpenAI’s third generation LLM, called GPT-3, started a race to build larger language models, with size measured by the number of model parameters and amount of training data. The race has since expanded to generative AI systems with text, images, videos, voice, and music as inputs and outputs (Fergusson, *et al.*, 2023). Prominent researchers have hailed these models as bringing us closer to AGI, with DeepMind senior director Nando de Freitas exclaiming that “it’s all about scale now! The Game is Over! It’s about making these models bigger ... solving these scaling challenges is what will deliver AGI” [107].

Unlike small, “narrow AI” models built for specific tasks and trained using curated datasets, systems that are advertised as having “broad general knowledge and domain expertise” require models with upwards of hundreds of billions of parameters and training datasets of upwards of hundreds of gigabytes [108]. The staggering environmental costs of training and performing inference on models of this size have been documented by a number of researchers (Luccioni, *et al.*, 2023). But from the TESCREAL perspective, this cost should not be of much concern, because the impending climate catastrophe does not pose an existential risk to humanity, while stopping work on building “value-aligned” AGI could (see Torres, 2022, 2021; Ord, 2020). Other AGI proponents, like DeepMind's Kavukcuoglu, promise that "advances in AGI research will supercharge society's ability to tackle and manage climate change," while the AGI race has been documented to do just the opposite [109].

In addition to these costs, the size of the datasets used in systems advertised to be stepping stones towards AGI also exacerbates the dangers caused by the lack of appropriate scoping discussed earlier, because model builders are less likely to curate, document and understand their datasets when they reach such sizes (Bender and Gebru, *et al.*, 2021). For instance, the LAION-5B dataset was taken down in December 2023 after Child Sexual Abuse Material (CSAM) was found in the dataset (Thiel, 2023; Cole, 2023; Birhane, *et al.*, 2021). The dataset was used to train models like Stability AI's Stable Diffusion which has millions of daily users (Jiang, *et al.*, 2023).

As detailed by a number of scholars, both the environmental impacts and the unsafe outputs of these systems disproportionately affect marginalized groups like racial and gender minorities, disabled people, and citizens of developing countries bearing the brunt of the climate catastrophe (Bender and Gebru, *et al.*, 2021). The AGI race not only perpetuates these harms to marginalized groups, but it does so while depleting resources from these same groups to pursue the race. Resources that could go to many entities around the world, each building computational systems that serve the needs of specific communities, are being siphoned away to a handful of corporations trying to build AGI. For instance, the CTO of Lesan AI, a machine translation startup specializing in a number of Ethiopian languages, reported that potential investors were discouraged from investing in his startup after believing that OpenAI and Meta had made his organization obsolete (Donastorg, 2023; Gebru, 2023), in spite of evidence demonstrating that some of their models perform poorly for the languages in question (Hadgu, *et al.*, 2023).

Thus, the end result of pursuing the AGI race has been an accumulation of resources by organizations like OpenAI (US$100B+ valuation) and Anthropic (US$18B+ valuation) that position themselves as leaders of an endeavor to "benefit all of humanity" (Tan, *et al.*, 2023; Field, 2023) [110], and a depletion of resources from the many organizations around the world working on tools to serve the needs of specific communities. Indeed, some leading TESCREALists have suggested that AGI should be developed by "some small vanguard of elite super-programmers and uber-scientists" (Goertzel, 2015), an attitude that mirrors that of first-wave eugenicists who used IQ tests to determine who is "fit" to lead society. Instead of having the multitudes of humans around the world building tools serving their own needs, the TESCREAList techno-utopia entails diverting resources to create their singular vision of a superior being with characteristics determined and controlled by them.

6.3. Evading accountability

The veneer of building a complex, all-knowing being, as imagined by TESCREALists, has given organizations cover to evade accountability for the labor exploitation and deceptive practices that, in practice, fuel the systems they advertise as stepping stones towards AGI. Organizations working on AGI depend on millions of exploited workers around the world (Gray and Suri, 2019; Williams, *et al.*, 2022) who label data to train, evaluate and moderate their systems. For example, in 2023, *Time* reported that Kenyan workers paid as low as US$1.32/hour were hired to label toxic content such as "textual descriptions of sexual abuse, hate speech, and violence" to help OpenAI develop automated filters that prevent the public from seeing these outputs (Perrigo, 2023). Workers also had to label images including those containing "bestiality, rape, and sexual slavery." In the process, they reported being "mentally scarred by the work," living in the opposite reality from an AGI ushered utopia where "people will be freed up to spend more time with people they care about" [111]. As Williams, *et al.* (2022) wrote, while corporations such as OpenAI headquartered in Silicon Valley receive billions of dollars in investment, with their executives and AI researchers paid six to seven figures, this salary is not afforded to the low-income essential workers around the world mitigating the harms of these systems at a cost to their own mental health.

Anthropomorphizing systems output by organizations striving to build AGI by calling them "thinking" or "sentient" machines obfuscates the many exploited humans involved in training and evaluating these systems, as well as the resources that are consumed in the process. Ensuring system safety requires an ecosystem of agencies, lawmakers, and internal groups that scrutinize organizational practices, audit processes by which products are built and deployed, and hold organizations accountable in cases of safety violations (Raji, *et al.*, 2020). When organizations advertise their systems as a step toward AGI, or when researchers ask if machines can learn "morality" and whether they "understand us" (Agüera y Arcas, 2022; Jiang, *et al.*, 2022), they move attention away from organizations' responsibility to create products with certain requirements, or protecting the wellbeing of the workers involved in the process, to discussions of AI systems as if they exist on their own (Tucker, 2022). This is particularly harmful because ascribing such agency to AI systems also misleads the public into the actual capabilities of these systems, which can result in erroneous or even harmful outcomes, while allowing organizations who build these products and encourage such uses to evade accountability (Gebru and Mitchell, 2022).

For example, OpenAI's leaders have described their tools as "slightly conscious" [112], and predict that "in the next five years, computer programs that can think will read legal documents and give medical advice" [113]. Venkatasubramanian discussed the anthropomorphization in ChatGPT's design in his interview with VentureBeat noting: "... Google Bard doesn't do this. Google Bard is a system for making queries and getting answers. ChatGPT puts little three dots [as if it's] 'thinking' just like your text message does. ChatGPT puts out words one at a time as if it's typing. The system is designed to make it look like there's a person at the other end of it. That is deceptive" (Goldman, 2023). In spite of the marketing of ChatGPT and similar systems as nearly all-knowing machines, they should not be used for search purposes for many reasons, one of them being that they can completely fabricate information while presenting it to users with confident-sounding prose (Bender and Gebru, *et al.*, 2021; Shah and Bender, 2022), resulting in what Bender has called the equivalent of an oil spill on the information ecosystem (Shah and Bender, 2024) [114].

However, due to the claims of those building these systems and the design choices that anthropomorphize them, organizations have used tools like ChatGPT in high-stakes scenarios like providing mental health advice without informing people that they were engaging with a chatbot (Biron, 2023), and claiming to be able to replace lawyers with chatbots fine-tuned on ChatGPT (Cerullo, 2023). In a tragic event, a man reportedly died by suicide in Belgium, after text output by an LLM based chatbot encouraged him to do so (Xiang, 2023). Nevertheless, OpenAI's former chief scientist recently tweeted that "in the future, once the robustness of our models will exceed some threshold, we will have *wildly effective* and dirt cheap AI therapy" [115]. Recently, a judge penalized two U.S. lawyers who used ChatGPT to generate fake court cases in their demand letters (Milmo and Agency, 2023). While OpenAI's terms of use simply place all responsibility to uphold the law on the user, thus evading accountability, its leaders simultaneously inspire such high-stakes uses of their products by associating their tools with AGI and implying that those capabilities are imminent.

A similar dynamic can be observed with text-to-image models like Stability AI's Stable Diffusion and OpenAI's Dall-E, which journalists have described as being "inspired" by artists, just like artists are inspired by other artists [116]. Jiang, *et al.* (2023) described the consequences of this type of anthropomorphization by noting that it "devalues artists' works, robs them of credit and compensation" for the data that is taken from them to train these models, "and ascribes accountability to the image generators rather than holding the entities that create them accountable." As Karla Ortiz remarked: "AI companies claimed to bring art to the masses, but ... they just gave potential art theft/plagiarism to the masses" [117]. By ascribing human-like qualities to models trained by corporations to generate profit, using troves of artists' works without obtaining their consent or compensating them, our attention is directed away from investigating the processes by which corporations create these products, the harms their practices cause artists, and the mechanisms that need to be put in place to hold the corporations accountable.

6.4. Co-opting safety

Another way in which those attempting to build AGI have evaded accountability is by framing the AGI race as existential to humans in spite of the harms caused by the race as detailed earlier. Like their first-wave eugenicist predecessors who believed that "improving the human stock" was the only way to safeguard "human civilization," leaders of the TESCREAL bundle argue that creating aligned AGI is a way to safeguard civilization, and thus, the most important task for humanity this century (Ord, 2020). As Yudkowsky remarked, "ours is the era of inadequate AI alignment theory. Any other facts about this era are relatively unimportant" [118].

Framing the AGI agenda as a safety issue allows companies working toward it to describe themselves as "AI safety" organizations safeguarding humanity's future, while simultaneously creating unsafe products, centralizing power, and evading accountability, as discussed earlier. According to some leading TESCREALists, such harms would be classified as nothing more than "mere ripples on the surface of the great sea of life," as Nick Bostrom (2009b) described the worst disasters and atrocities of the twentieth century. To them, the far more pressing "risk" arises from the possibility of never realizing the eugenic ideals promised through creating "transhuman AGI" (Goertzel, 2010) that is orders of magnitude more "intelligent" and "morally superior" (Fitzgerald, *et al.*, 2020) than human beings. TESCREALists Greaves and MacAskill (2019) write that "for the purposes of evaluating actions, we can in the first instance often simply ignore all the effects contained in the first 100 (or even 1,000) years, focussing primarily on the further-future effects. Short-run effects act as little more than tie-breakers."

TESCREAList leaders have argued that AGI is inevitable — someone is going to build it [119]. If it is built by those who are not "value-aligned," and according to them China would fit into this category, it would be a national security risk to those Western countries that TESCREAL leaders are from, or worse, could render all of humanity extinct (Davis, 2023). Hence, Western nations should make building "value-aligned" AGI national priorities, as they are the ones who can build "value-aligned" AGI beneficial to "all of humanity" (Davis, 2023). By tapping into Cold War rhetoric and framing the need to build AGI as a safety concern, TESCREALists have started to steer Western politicians and global multilateral organizations into investing in, legitimizing, and prioritizing their AGI agenda. Weiss-Blatt has detailed the amount of lobbying and media influence campaigns by TESCREAL organizations to ensure that the hypothetical AGI apocalypse is prioritized by policy-makers worldwide [120].

This legitimization has had concrete policy impacts, with legislators who may not want to advance TESCREAList utopian ideals nonetheless being heavily influenced by them. An investigation by *Politico* detailed how this influence is steering U.S. and U.K. AI policy towards preventing a hypothetical human extinction event from nonexistent superintelligent machines (Bordelon, 2023; Clarke, 2023). Meanwhile, corporations like OpenAI are evading scrutiny by presenting their products as too powerful for regulators to understand and regulate, while exploiting labor and profiting off of people's data without consent or compensation as noted earlier. In May 2023, Sam Altman testified before the U.S. Senate urging regulation on AI, and warned in a blogpost less than a week later that there needs to be an agency, akin to the International Atomic Agency, to regulate "superintelligence" since it "will be more powerful than other technologies humanity has had to contend with in the past" [121]. He was shortly after described in the media as an "Oppenheimer of our age" warning about his own powerful creation (Weil, 2023). However, while warning the public about the dangers of nonexistent superintelligent machines that would need to be regulated, OpenAI was simultaneously threatening to exit the EU, stating that the draft of the EU AI Act at the time would result in "over-regulating" them (Reuters, 2023a).

In this way, TESCREALists have been able to divert resources toward trying to build AGI and stopping their version of an apocalypse in the far future, while dissuading the public from scrutinizing the actual harms that they cause in their attempts to build AGI. As another example, Max Tegmark, cofounder of the Future of Life Institute along with Jaan Tallinn, delivered a talk at the 2017 Effective Altruism conference (EA Global) in which he argued that "if we don't improve our technology, we are doomed ... but with tech, life can flourish for billions of years" [122]. But in 2023 the Future of Life Institute circulated a widely publicized petition signed by Tegmark and many of those responsible for the AGI race, including Musk and Altman, to "pause giant AI experiments" to stop "nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us" [123]. Appearing on *Democracy Now!*, when asked about the present day dangers like biometric surveillance warned by researchers and activists like Tawnana Petty, Tegmark stated: “Extinction is not something in the very distant future ... And once we’re all extinct, you know, all these other issues cease to even matter” (Bengio, *et al.*, 2023). Similarly, when Geoffery Hinton, who also signed the petition, was asked by *Rolling Stone* about the issues raised by Timnit Gebru who was fired by Google after writing a paper on the dangers of large language models, he answered: “I believe that the possibility that digital intelligence will become much smarter than humans and will replace us as the apex intelligence is a more serious threat to humanity than bias and discrimination” (O’Neil, 2023).

While ringing the alarm about these hypothetical apocalyptic scenarios, TESCREALists simultaneously create organizations to pursue building the very AGI that they warn could render us extinct, because they would do it in a way that is “safe” and “beneficial” as noted by companies like OpenAI, DeepMind, Anthropic, and Musk’s xAI. Three months after signing the “Pause AI” letter, Elon Musk announced his new AI organization aiming to develop AGI that is “maximally curious” (Reuters, 2023b). He was joined, in an advisory role, by Dan Hendrycks, founder of the Center of AI Safety, a TESCREAList institute which circulated a 22-word statement similar to the “pause letter,” warning about the existential risks of AI [124]. Thus, TESCREALists use the language of “safety” to first drive resources into the goal of building AGI, which in turn causes harm to marginalized groups, and then use the language of “safety” to dissuade investigations into those harms and once again divert resources into preventing a hypothetical AGI apocalypse.

7. Building well-scoped and well-defined systems

While those working to build AGI describe their work as scientific and engineering endeavors, we argue that attempting to build AGI follows neither scientific nor engineering principles. The scientific method often involves postulating specific hypotheses and testing them with extensive experimentation (Hepburn and Andersen, 2021). Engineering requires us to provide specifications for expected behavior, tolerance, and safety protocols for the tools that we build (Khlaaf, 2023; Kossiakoff, *et al.*, 2020). Engineers often model idealized versions of their systems, as well as nonidealities and their impacts on system functionality (Khlaaf. 2023; Kossiakoff, *et al.*, 2020; Tripathy and Naik, 2011). They then perform stress tests to understand the behavior of the systems they build under various circumstances: those considered standard operating conditions, and those deviating from the norm (Kossiakoff, *et al.*, 2020).

As an example, one of us worked as a hardware engineer designing audio circuitry for devices such as laptops. Some of the tests that we performed as part of our work included drop testing, constantly dropping devices to understand the manner in which their functionality degrades when they are exposed to shocks [125], placing the devices in extremely cold or hot environments (International Electrotechnical Commission, 2020), frequently restarting them [126], and performing different types of tests to understand the behavior of these systems under conditions that they were not normally meant to operate in. These stress tests occurred in addition to extensive testing and documentation under conditions that the devices were meant to be operational. Engineers had to ensure compliance with various laws such as the Restriction of Hazardous Substances Directive in the European Union, which requires electrical and electronic components to be free of hazardous materials such as lead (European Union, 2011).

What would be the standard operating conditions for a system advertised as a “universal algorithm for learning and acting in any environment” [127]? How could experiments designed to test the functionality of such a system have construct validity: the ability of an experiment to faithfully portray a system’s expected performance in the real world (O’Leary-Kelly and Vokurka, 1998)? We argue that these are not questions that can be answered for a system like AGI, which is not well defined but is purported to be able to accomplish infinitely many tasks under infinitely many conditions. Hence, while TESCREALists like Goertzel lamented the focus on “narrow AI” described as “collections of dumb specialists in small domains” before the current resurgence of the excitement towards AGI [128], we argue that the first steps in making any AI system safe are attempting to build well-scoped and well-defined systems like those described as “narrow AI,” rather than a machine that can supposedly do any task under any circumstance.

The AGI race is not an inevitable, unstoppable march towards technological progress, grounded in careful scientific and engineering principles (van Rooij, *et al.*, 2023). It is a movement created by adherents of the TESCREAL bundle seeking to “safeguard humanity” (Ord, 2020) by, in Altman’s words, building a “magic intelligence in the sky” (Germain, 2023), just like their first-wave eugenicist predecessors who thought they could “perfect” the “human-stock” through selective breeding (see Bloomfield, 1949). Through concerted campaigns to influence AI research and policy practices backed by billions of dollars, TESCREALists have steered the field into prioritizing attempts to build unscoped systems which are inherently unsafe, and have resulted in documented harms to marginalized groups. As reported by Nitasha Tiku (2023), Open Philanthropy alone has spent more than half a billion dollars on initiatives pertaining to AGI. They have done this by:

> developing a pipeline of talent to fight rogue AI, building a scaffolding of think tanks, YouTube channels, prize competitions, grants, research funding and scholarships — as well as a new fellowship that can pay student leaders as much as $80,000 a year, plus tens of thousands of dollars in expenses.

This investment has succeeded in legitimizing the AGI race such that many students and practitioners who may not be aligned with TESCREAL utopian ideals are working to advance the AGI agenda because it is presented as a natural progression in the field of AI. In the same way that first-wave eugenicists and race scientists sought and achieved academic legitimacy for their research (Saini, 2019), TESCREALists have created a veneer of scientific authority that makes their ideas more palatable to uncritical audiences, and thus have succeeded in influencing research and policy directions in the field of AI. First-wave eugenics proved to be ineffective and catastrophic. But as Jean Gayon and Daniel Jacobi signify with the term “eternal return of eugenics,” eugenic ideals keep on being repackaged in different forms [129]. The AGI race is yet another attempt, diverting resources and attention away from potentially useful research directions, and causing harm in the process of trying to achieve a techno-utopian ideal crafted by self appointed “vanguards” of humanity.

8. Conclusion

In this paper, we have asked: what motivates those who aspire to build AGI — a system, which, although it does not have one definition even among those who claim to be building it, seems to be an all-knowing machine akin to a “god”? The answer is that the current push for AGI is driven by a set of ideologies which we label the “second wave” of eugenics. Leaders of the AGI movement subscribe to this set of ideologies, which directly emerged from the modern eugenics movement, and therefore have inherited similar ideals. We trace the influence of this set of ideologies throughout the AGI movement, and show the manner in which its harmful ideals have resulted in systems that perpetuate inequity, centralize power, and harm the same groups that were targeted by the first-wave modern eugenics movement. We argue that attempting to build something akin to a god is an inherently unsafe practice, and urge researchers and practitioners to abandon this goal in lieu of building well-defined, well-scoped systems that prioritize people’s safety.

About the authors

Timnit Gebru is founder and executive director of the Distributed Artificial Intelligence Research Institute (DAIR). Her research interests include addressing the harms of AI systems towards marginalized groups

and creating community-rooted AI systems.
E-mail: timnit [at] dair-institute [dot] org

Émile P. Torres is a postdoctoral scholar at the Inamori International Center for Ethics and Excellence at Case Western Reserve University. Their research focuses on the ethics of AI and the history of thinking about human extinction within the Western tradition.
E-mail: philosophytorres [at] gmail [dot] com

Acknowledgements

We thank Alexander Thomas, Beth Singler, Douglas Rushkoff, Emily M. Bender, Jacob Metclaf, Jenna Burrell, Katherine Heller, Keira Havens, Remmelt Ellen, Samy Bengio, Syed Mustafa Ali, Nicholas Rodelo, and the anonymous reviewers for *First Monday* and FAccT for reviewing drafts of this paper and providing thorough feedback that has allowed us to make it stronger.

Notes

[1] "OpenAI charter," at https://openai.com/charter, accessed 30 January 2024.

[2] Pennachin and Goertzel, 2007a, p. 1.

[3] "What is AGI?" at https://medium.com/intuitionmachine/what-is- agi-99cdb671c88e, accessed 30 January 2024.

[4] Russell and Norvig, 2010, p. 27.

[5] Markoff, 2005; Russell and Norvig, 2010, pp. 16–28.

[6] Russell and Norvig, 2010, pp. 16–28; Pennachin and Goertzel, 2007a.

[7] Note that we have capitalized some of the TESCREAL ideologies but not others. We are following what members of these ideologies themselves capitalize: Extropians, Rationalists, and Effective Altruists prefer to capitalize these terms, whereas transhumanists, singularitarians, cosmists, and longtermists do not.

[8] Goertzel, 2010, p. 231.

[9] Sam Altman blog, "Planning for AGI and beyond " (23 February), at https://openai.com/blog/planning-for-agi-and-beyond, accessed 30 January 2024.

[10] Preface, in Pennachin and Goertzel (2007b).

[11] Note that our discussion of eugenics is incomplete, due largely to space limitations. In what follows, we focus on providing a brief background to situate the emergence of the TESCREAL bundle of ideologies.

[12] Gaca, 2003, p. 52.

[13] Galton, 1998, p. 265.

[14] Fogarty and Osborne, 2010, p. 334.

[15] Galton in Zuberi, 2001, p. 43.

[16] For an insightful discussion of the history of eugenics in California, see Harris (2023).

[17] Bashford and Levine, 2010, p. 13.

[18] See Harris (2023) for further discussion.

[19] History page of the Adelphi Genetic Forum, retrieved on January 31, 2024 from https://adelphigenetics.org/history/. They also write that "despite our organisation's former name, the Adelphi Genetics Forum rejects outright the theoretical basis and practice of coercive eugenics, which it regards as having no place in modern life."

[20] Quine, 2010, p. 392.

[21] Alternatively, such technologies, according to these eugenicists, could potentially enable individuals to radically modify themselves within a single generation (see Bostrom, 2013).

[22] See section 5 of Bostrom (2005c).

[23] "Transhumanism. What it is. What it is not," at https://www.youtube.com/watch?v=sdjMoykqxys&t=425s, accessed 31 January 2024.

[24] See "Eliezer Yudkowsky," at https://extropians.weidai.com/extropians.96/author.html#441, accessed 31 January 2024.

[25] "Welcome to LessWrong!" (14 June 2019), at https://www.lesswrong.com/posts/bJ2haLkcGeLtTWaD5/welcome-to-lesswrong, accessed 31 January 2024.

[26] "Top Contributors," *Machine Intelligence Research Institute*, at https://intelligence.org/topcontributors/, accessed 30 January 2024.

[27] "Top Contributors," *Machine Intelligence Research Institute*, at https://intelligence.org/topcontributors/, accessed 30 January 2024. See also 'Peter Thiel's keynote — Effective Altruism Summit 2013," at https://www.youtube.com/watch?v=h8KkXcBwHec&t=818s&ab_channel=nnevvinn, accessed 2 February 2024; and Bohan (2022), pp. 44–53.

[28] "Rationalist Movement," at https://www.lesswrong.com/tag/rationalist-movement, accessed 31 January 2024.

[29] "Top Contributors," *Machine Intelligence Research Institute*, at https://intelligence.org/topcontributors/, accessed 30 January 2024.

[30] "Donation Advice," *EA Funds*, at https://funds.effectivealtruism.org/donation-advice, accessed 17 January 2024.

[31] "Research staff," *Future of Humanity Institute*, at https://web.archive.org/web/20070209123709/http://www.fhi.ox.ac.uk:80/staff.html, accessed 15 January 2024.

[32] "Dustin Moskovitz," *Effective Altruism Forum*, at https://forum.effectivealtruism.org/topics/dustin-moskovitz, accessed 31 January 2024.

[33] "Sam Bankman-Fried," *Effective Altruism Forum*, at https://forum.effectivealtruism.org/topics/sam-bankman-fried#:~:text=Before%20the%20FTX%20collapse%2C%20Bankman,his%20wealth%20to%20longtermist%20causes, accessed 31 January 2024.

[34] Huxley, 1957, p. 17.

[35] “The extropian principles,” at https://www.mrob.com/pub/religion/extro_prin.html, accessed 31 January 2024.

[36] Bostrom, 2005a, p. 12.

[37] “Re: SOCIETY: The Quiet Revolution.” at https://extropians.weidai.com/extropians.96/4858.html, accessed 31 January 2024.

[38] See “Neologisms of Extropy,” at https://web.archive.org/web/20060220131448/https://www.extropy.org/neologo.htm, accessed 1 February 2024; and “The Singularitarian Principles,” at https://web.archive.org/web/20120403111339/http://yudkowsky.net/obsolete/principles.html, accessed 31 January 2024.

[39] “LessWrong comment,” at https://www.lesswrong.com/posts/aFtWRL3QihoF5uQd5/guardians-of-the-gene- pool?commentId=BzFBAQhRRyMCk7Wny, accessed 31 January 2024; and “Neologisms of Extropy,” at https://web.archive.org/web/20060220131448/https://www.extropy.org/neologo.htm, accessed 1 February 2024.

[40] “About SingularityNET,” at https://singularitynet.io/aboutus/, accessed 14 January 2024.

[41] Goertzel, 2010, p. 10.

[42] Historically, “cosmism” can be traced back to the latter nineteenth century work of Nikolai Fyodorov, which was later developed by Russian scientists like Konstantin Tsiolkovsky and Vladimir Vernadsky (Young, 2012). Russian cosmism can be seen as a precursor to modern transhumanism, although the particular version that we are interested in specifically arises from the work of Goertzel, which is distinct from the “cosmism” of earlier Russian theorists.

[43] “Welcome to LessWrong!” at https://www.lesswrong.com/posts/bJ2haLkcGeLtTWaD5/welcome-to-lesswrong, accessed 31 January 2024.

[44] “Singularity,” at https://www.lesswrong.com/tag/singularity, accessed 31 January 2024; See also “Transhumanism as Simplified Humanism,” at https://www.lesswrong.com/posts/Aud7CL7uhz55KL8jG/transhumanism-as-simplified-humanism, accessed 31 January 2024.

[45] Centre of Effective Altruism website, at https://www.centreforeffectivealtruism.org/, accessed 30 January 2024.

[46] Beckstead, 2013, p. ii.

[47] It is unclear to us what methodology Bostrom used to arrive at such astronomical numbers. Nonetheless, he writes that what matters for longtermism “is not the exact numbers but the fact that they are huge” (Bostrom, 2003; for criticism of such calculations, see Torres, 2024).

[48] For example, a 2019 survey of the EA community found that “a clear majority of EAs (80.7 percent) identified with consequentialism, especially utilitarian consequentialism” (“EA Survey 2019: Community Demographics & Characteristics,” at https://rethinkpriorities.org/publications/eas2019-community-demographics-characteristics, accessed 31 January 2024). The central thrust of Bostrom’s (2003) influential paper comes from totalist utilitarianism, and this paper is considered to be one of the founding documents of longtermism (see, *e.g.*, footnote 27 of chapter 2 in Ord, 2020; Greaves and MacAskill, 2019, p. 3). Finally, in a keynote address at the Effective Altruism Global 2016 conference, Ord explicitly argued that

"core ideas such as the Scientific Revolution, the Enlightenment and Utilitarianism have greatly contributed to the upbringing of effective altruism" (at https://www.youtube.com/watch?v=VH2LhSod1M4&ab_channel=CentreforEffectiveAltruism, accessed 31 January 2024).

49. Note that EA also has roots going back to the "global ethics" of the utilitarian Peter Singer, who also advocated for infanticide of disabled children, writing "we think that some infants with severe disabilities should be killed" (quoted in Ekland-Olson, 2011, p. 204). However, the EA movement was cofounded circa 2009 by Toby Ord, who coauthored an article with Bostrom three years earlier that essentially defended the transhumanist position on human enhancement (Bostrom and Ord, 2006). Ord was, furthermore, a Research Associate at Bostrom's Future of Humanity Institute (FHI) as early as 2007, and over the past several years, the EA movement as a whole has been shifting toward longtermist considerations (Matthews, 2022), some of which are more or less overtly transhumanist (see, *e.g.*, Ord's (2020) discussion of "our potential" in the section titled "Quality" of *The Precipice*).

50. Sam Altman's Twitter post (14 August 2022), at https://twitter.com/sama/status/1559011065899282432?lang=en, accessed 31 January 2024.

51. Elon Musk's Twitter post (2 August 2022), at https://twitter.com/elonmusk/status/1554335028313718784, accessed 30 January 2024.

52. Elon Musk's Twitter post (2 May 2023), at https://twitter.com/elonmusk/status/1653421967570096128, accessed 30 January 2024. Many accounts of Musk's worldview are consistent with our claim that Musk is a TESCREAList. For example, Ashlee Vance wrote that Musk is "a card-carrying member of Silicon Valley's techno-utopian club," which anticipates that "one day, soon enough, we'll be able to download our brains to a computer, relax, and let their algorithms take care of everything. ... More disconcerting is their underlying message that humans are flawed and our humanity is an annoying burden that needs to be dealt with in due course." Vance added that Musk's "highfalutin talk often sounded straight out of the techno-utopian playbook" (Vance, 2016). Note also that our notion of TESCREALism is very close to what Douglas Rushkoff (2022) calls "The Mindset."

53. Quoted in Parrinder, 1997, p. 2. See also Francis Galton, 2001. "The Eugenic College of Kantsaywhere," Critical edition, transcribed and edited by Lyman Tower Sargent, *Utopian Studies*, volume 12, number 2, pp. 191–209.

54. Kurzweil, 1999, p. 141.

55. See "Why AI will save the world" (6 June 2023), at https://a16z.com/ai-will-save-the-world/, accessed 31 January 2024; "Planning for AGI and beyond" (24 February 2023), at https://openai.com/blog/planning-for-agi-and-beyond, accessed 31 January 2024; and "Let's think about slowing down AI" (22 December 2022), at https://worldspiritsockpuppet.substack.com/p/lets-think-about-slowing-down-ai, accessed 31 January 2024.

56. "Apology for an Old Email," at https://nickbostrom.com/oldemail.pdf, accessed 31 January 2024.

57. "Re: Profiting on tragedy? (was Humour)," at https://diyhpl.us/~bryan/irc/extropians/extracted-extropians-archive/archive/9612/4881.html, accessed 31 January 2024.

58. "Adamantly preventing tradgedy? (was Humour)," at https://extropians.weidai.com/extropians.96/4759.html, accessed 31 January 2024.

59. Eliezer Yudkowsky Twitter post, archived on 13 February 2023, at https://web.archive.org/web/20230213224423/https://twitter.com/esyudkowsky/status/1131208777032560640, accessed 1 February 2024.

60. Bio in Marc Andreessen's Twitter biography, archived on 23 May 2023, at

https://web.archive.org/web/20230523005947/https://twitter.com/pmarca, accessed 31 January 2024.

61. "How dependent is the effective altruism movement on Dustin Moskovitz and Cari Tuna?" (21 September 2020), at https://forum.effectivealtruism.org/posts/4BJSXH9ho4eYNT73P/how-dependent-is-the- effective-altruism-movement-on-dustin, accessed 31 January 2024.

62. Marc Andreessen, "The techno-optimist manifesto" (16 October 2023), at https://a16z.com/the-techno-optimist-manifesto/, accessed 31 January 2024.

63. See Marc Andreessen's Twitter handle, at https://twitter.com/pmarca, accessed 31 January 2024.

64. See Garry Tan's Twitter biography, at https://twitter.com/garrytan, accessed 30 January 2024.

65. McCarthy, *et al.*, 1955; Russell and Norvig, 2010, pp. 16–28.

66. Markoff, 2005; Russell and Norvig, 2010, pp. 16–28.

67. Preface, in Pennachin and Goertzel (2007b).

68. Voss, 2007, p. 153.

69. Blog post by Ben Goertzel: "Who coined the term 'AGI'?" (28 August 2011), at https://goertzel.org/who-coined-the-term-agi/, accessed 31 January 2024.

70. Preface, in Pennachin and Goertzel (2007b).

71. "AGI Researchers Stop Quoting White Supremacists Challenge (Impossible)," *Medium* (22 September 2023), at https://medium.com/@collegehill/agi-researchers-stop-quoting-white-supremacists-challenge-impossible-d1002469d572, accessed 31 January 2024.

72. Gottfredson, 1997, p. 91.

73. Southern Poverty Law Center on Linda Gottfredson, at https://www.splcenter.org/fighting-hate/extremist-files/individual/linda-gottfredson, accessed 31 January 2024.

74. Keira Haven's Twitter post (22 September 2023), at https://twitter.com/Keira_Havens/status/1705404087276372121, accessed 30 January 2024.

75. Archived Web page of Artificial General Intelligence Research Institute, at https://web.archive.org/web/20080512012745/http://www.agiri.org/wiki/Main_Page, accessed 31 January 2024.

76. This was written by Bruce Klein on OpenCog (https://wiki.opencog.org/wikihome/index.php?title=Artificial_General_Intelligence&oldid=1297, accessed 31 January 2024). Klein lists a novamente.net e-mail address (https://wiki.opencog.org/w/User:Bruceklein, accessed 31 January 2024). Novamente is an AI company founded by Goertzel (https://www.crunchbase.com/organization/novamente-llc, accessed 31 January 2024).

77. "Top Contributors," *Machine Intelligence Research Institute*, at https://intelligence.org/topcontributors/, accessed 30 January 2024.

78. "About MIRI," at https://intelligence.org/about/, accessed 30 January 2024. Note that the Singularity Institute's original mission was explicitly accelerationist: "to accelerate toward artificial intelligence" (Torres, 2023c).

79. DeepMind AGI, DeepMind's website before merging with Google Brain in 2023 (see announcement

from Google CEO (20 April 2023) at https://blog.google/technology/ai/april-ai-update/, accessed 31 January 2024), from http://deepmindagi.com/.

[80] Video of Future of Life Institute panel on “Superintelligence: Science or fiction?” at https://www.youtube.com/watch?v=h0962biiZa4, accessed 31 January 2024.

[81] “Measuring machine intelligence — Shane Legg, Singularity Summit 2010,” at https://www.youtube.com/watch?v=0ghzG14dT-w, accessed 31 January 2024.

[82] “Singularity Summit: An Annual Conference on Science, Technology, and the Future,” at https://intelligence.org/singularitysummit/, accessed 31 January 2024.

[83] “Why we launched DeepMind Ethics & Society” (3 October 2017), at https://deepmind.google/discover/blog/why-we-launched-deepmind-ethics-society/, accessed 31 January 2024.

[84] “OpenAI Charter” (9 April 2018), at https://openai.com/charter, accessed 31 January 2024.

[85] Open Philanthropy, “OpenAI — General Support,” at https://www.openphilanthropy.org/grants/openai-general-support/, accessed 31 January 2024.

[86] “OpenAI LP,” at https://openai.com/blog/openai-lp, accessed 31 January 2024.

[87] Microsoft announcement: “OpenAI forms exclusive computing partnership with Microsoft to build new Azure AI supercomputing technologies” (22 July 2019), at https://news.microsoft.com/2019/07/22/openai-forms-exclusive-computing-partnership- with-microsoft-to-build-new-azure-ai-supercomputing-technologies/, accessed 31 January 2024.

[88] Antrhopic’s website, at https://www.anthropic.com/company, accessed 31 January 2024.

[89] “The highest-impact career paths our research has identified so far,” at https://80000hours.org/career-reviews/, accessed 31 January 2024.

[90] “About xAI,” at https://www.x.ai/about/, accessed 31 January 2024.

[91] Open Philanthropy, “Center for AI Safety — General Support (2022),” at https://www.openphilanthropy.org/grants/center-for-ai-safety-general-support/, accessed 31 January 2024.

[92] “Introduction to Pragmatic AI Safety” (9 May 2022), at https://forum.effectivealtruism.org/posts/MskKEsj8nWREoMjQK/introduction-to-pragmatic-ai-safety-pragmatic-ai-safety-1, accessed 31 January 2024.

[93] “Expanding access to safer AI with Amazon” (25 September 2023), at https://www.anthropic.com/news/anthropic-amazon, accessed 31 January 2024.

[94] Preface, in Pennachin and Goertzel (2007b).

[95] “Real-world challenges for AGI” (2 November 2021), at https://deepmind.google/discover/blog/real-world-challenges-for-agi/, accessed 31 January 2024.

[96] Goertzel, 2010, p. 230.

[97] Goertzel, 2010, p.11.

[98] Sam Altman’s Twitter post, at https://twitter.com/sama/status/1520798948562141184, accessed 31 January 2024.

99. “Moore’s Law for Everything” (16 March 2021), at https://moores.samaltman.com/, accessed 31 January 2024. Incidentally, the idea that the Singularity is “unstoppable” or “inevitable” is strongly emphasized in Kurzweil’s work, as when he writes that “the Singularity denotes an event that will take place in the material world, the inevitable next step in the evolutionary process that started with biological evolution and has extended through human-directed technological evolution” (Kurzweil, 2005).

100. “Governance of superintelligence” (22 May 2023), at https://openai.com/blog/governance-of-superintelligence, accessed 31 January 2024.

101. “Introducing Superalignment” (5 July 2023), at https://openai.com/blog/introducing-superalignment, accessed 31 January 2024.

102. See section 1 of Torres (2019) for a list of such estimates; see also the TESCREAList Paul Christiano’s estimate that the “probability that humanity has somehow irreversibly messed up our future within 10 years of building powerful AI [is] 46%,” at https://www.lesswrong.com/posts/xWMqsvHapP3nwdSW8/my-views-on-doom, accessed 31 January 2024.

103. OpenAI Gym Beta, released in 2016, described as “a toolkit for developing and comparing reinforcement learning (RL) algorithms,” at https://openai.com/research/openai-gym-beta, accessed 31 January 2024.

104. “Transforming work and creativity with AI,” at https://openai.com/product, accessed 31 January 2024.

105. “Twitter post announcing Galactica using the language we described,” at https://twitter.com/paperswithcode/status/1592546933679476736, accessed 31 January 2024.

106. Website of Lesan AI, a machine translation company focused on Ethiopian languages, at https://lesan.ai/about.html, accessed 31 January 2024.

107. Nando de Freitas’ Twitter post, at https://twitter.com/NandoDF/status/1525397036325019649, accessed 31 January 2024.

108. “Model size and performance” section, part of Scale’s “Guide to large language models,” at https://scale.com/guides/large-language-models#model-size-and-performance, accessed 31 January 2024.

109. “Real-world challenges for AGI” (2 November 2021), at https://deepmind.google/discover/blog/real-world-challenges-for-agi/, accessed 31 January 2024.

110. “Planning for AGI and beyond” (24 February 2023), at https://openai.com/blog/planning-for-agi-and-beyond, accessed 31 January 2024. “Core views on AI safety: When, why, what, and how” (8 March 2023), at https://www.anthropic.com/news/core-views-on-ai-safety, accessed 31 January 2024.

111. Sam Altman, “Moore’s law for everything” (16 March 2021), at https://moores.samaltman.com/, accessed 31 January 2024.

112. Twitter post by Ilya Sutskever, OpenAI’s former chief scientist,at https://twitter.com/ilyasut/status/1491554478243258368, accessed 31 January 2024.

113. Sam Altman, “Moore’s law for everything” (16 March 2021), at https://moores.samaltman.com/, accessed 31 January 2024.

114. Emily M. Bender, “‘Ensuring safe, secure, and trustworthy AI’: What those seven companies avoided committing to,” *Medium* (29 July 2023), at https://medium.com/@emilymenonbender/ensuring-safe-secure-and-trustworthy-ai-what-those- seven-companies-avoided-committing-to-8c297f9d71a, accessed 31 January 2024.

115. Twitter post by Ilya Sutskever, OpenAI's former chief scientist, at https://twitter.com/ilyasut/status/1707027536150929689, accessed 31 January 2024.

116. Karla Ortiz, "Why AI Models are not inspired like humans" (7 December), at https://www.kortizblog.com/blog/why-ai-models-are-not-inspired-like-humans, accessed 31 January 2024.

117. *Ibid.*

118. Yudkowsky's Twitter profile, archived 14 March 2021, at https://web.archive.org/web/20210314211620/https://twitter.com/ESYudkowsky?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor, accessed 31 January 2024.

119. Kurzweil, 2005; Goertzel, 2010, p. 227; Bostrom, 2009a; Shin, 2023. Benjamin Hilton, "Preventing an AI-related catastrophe," at https://80000hours.org/problem-profiles/artificial-intelligence/, accessed 31 January 2024. See also Scott Alexander, "Why not slow AI progress?" (8 August 2022), at https://www.astralcodexten.com/p/why-not-slow-ai-progress, accessed 31 January 2024.

120. "The AI Panic Campaign — part 1" (15 October 2023), at https://www.aipanic.news/p/the-ai-panic-campaign-part-1, accessed 31 January 2024; and "The AI Panic Campaign — part 2" (15 October 2023), at https://www.aipanic.news/p/the-ai-panic-campaign-part-2, accessed 31 January 2024.

121. "Written Testimony of Sam Altman Chief Executive Officer OpenAI Before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, & the Law," at https://www.judiciary.senate.gov/imo/media/doc/2023-05-16%20-%20Bio%20&%20Testimony%20-%20Altman.pdf, accessed 31 January 2024. "Governance of superintelligence," *OpenAI* (22 May 2023), at https://openai.com/blog/governance-of-superintelligence, accessed 31 January 2024.

122. Tegmark, 2017, 2:30 timestamp.

123. "Pause Giant AI Experiments: An Open Letter" (22 March 2023), at https://futureoflife.org/open-letter/pause-giant-ai-experiments/, accessed 31 January 2024.

124. "Statement on AI Risk," at https://www.safe.ai/statement-on-ai-risk, accessed 31 January 2024.

125. "Board Level Drop Test Method of Components for Handheld Electronic Products" (JESD22-B111A), at https://www.jedec.org/standards-documents/docs/jesd-22-b111, accessed 1 February 2024.

126. "Reboot Tests (Device Fundamentals)," at https://learn.microsoft.com/en-us/windows-hardware/drivers/devtest/reboot-tests--device-fundamentals-, accessed 1 February 2024.

127. Russell and Norvig, 2010, p. 27.

128. Minsky in Preface, Pennachin and Goertzel (2007b).

129. Jean Gayon and Daniel Jacobi in Turda, 2010, p. 64.

References

Nicholas Agar, 1998. "Liberal eugenics," *Public Affairs Quarterly*, volume 12, number 2, pp. 137–155.

Blaise Agüera y Arcas, 2022. "Do large language models understand us?" *Daedalus*, volume 151, number 2, pp. 183–197.
doi: https://doi.org/10.1162/daed_a_01909, accessed 10 March 2024.

Blaise Agüera y Arcas and Peter Norvig, 2023. "Artificial general intelligence is already here," *Noema* (10 October), at https://www.noemamag.com/artificial-general-intelligence-is-already-here/, accessed 31 January 2024.

Shazeda Ahmed, Klaudia Jazwinska, Archana Ahlawat, Amy Winecoff, and Mona Wang, 2024. "Field-building and the epistemic culture of AI safety," *First Monday*, volume 29, number 4.
doi: https://dx.doi.org/10.5210/fm.v29i4.13626, accessed 10 March 2024.

Syed Mustafa Ali, 2019. "'White crisis' and/as 'existential risk,' the entangled apocalypticism of artificial intelligence," *Zygon*, volume 54, number 1, pp. 207–224.
doi: https://doi.org/10.1111/zygo.12498, accessed 10 March 2024.

Alison Bashford and Philippa Levine, 2010. "Introduction: Eugenics and the modern world," In: Alison Bashford and Philippa Levine (editors). *Oxford handbook of the history of eugenics*. Oxford: Oxford University Press, pp. 2–24.
doi: https://doi.org/10.1093/oxfordhb/9780195373141.013.0001, accessed 10 March 2024.

Dina Bass, 2023. "Microsoft invests $10 billion in ChatGPT maker OpenAI," *Bloomberg* (23 January), at https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai, accessed 31 January 2024.

Nicholas Beckstead, 2013. "On the overwhelming importance of shaping the far future," Ph.D. dissertation, Rutgers University, at https://rucore.libraries.rutgers.edu/rutgers-lib/40469/PDF/1/play/, accessed 10 March 2024.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, 2021. "On the dangers of stochastic parrots: Can language models be too big?" *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623.
doi: https://doi.org/10.1145/3442188.3445922, accessed 10 March 2024.

Yoshua Bengio, Max Tegmark, and Tawana Petty, 2023. "Artificial intelligence 'godfathers' call for regulation as rights groups warn AI encodes oppression," *Democracy Now!* (1 June), at https://www.democracynow.org/2023/6/1/ai_bengio_petty_tegmark, accessed 31 January 2024.

Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe, 2021. "Multimodal datasets: Misogyny, pornography, and malignant stereotypes," *arXiv*:2110.01963 (5 October).
doi: https://doi.org/10.48550/arXiv.2110.01963, accessed 31 January 2024.

Bethany Biron, 2023. "Online mental health company uses ChatGPT to help respond to users in experiment — raising ethical concerns around healthcare and AI technology," *Business Insider* (7 January), at https://www.businessinsider.com/company-using-chatgpt-mental-health-support-ethical-issues-2023-1, accessed 31 January 2024.

Edwin Black, 2003. "The horrifying American roots of Nazi eugenics," *History News Network*, at https://historynewsnetwork.org/article/1796, accessed 31 January 2024.

Lucy Bland and Lesley Hall, 2010. "Eugenics in Britain: The view from the metropole," In: Alison Bashford and Philippa Levine (editors). Oxford handbook of the history of eugenics. Oxford: Oxford University Press, pp. 212–227.
doi: https://doi.org/10.1093/oxfordhb/9780195373141.013.0012, accessed 10 March 2024.

Paul Bloomfield, 1949. "The eugenics of the Utopians: The utopia of the eugenists," *Eugenics Review*, volume 40, number 4, pp. 191–198.

Elise Bohan, 2022. *Future superhuman: Our transhuman lives in a make-or-break century*. Sydney, NSW: NewSouth Publishing.

Brendan Bordelon, 2023. “How a billionaire-backed network of AI advisers took over Washington,” *Politico* (13 October), at https://www.politico.com/news/2023/10/13/open-philanthropy-funding-ai-policy-00121362, accessed 1 February 2024.

Nick Bostrom, 2014. *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

Nick Bostrom, 2013. “Why I want to be a posthuman when I grow up,” In: Ronald L. Sandler (editor). *Ethics and emerging technologies*. London: Palgrave Macmillan, pp. 218–234.
doi: https://doi.org/10.1057/9781137349088_15, accessed 10 March 2024.

Nick Bostrom, 2009a. “The future of humanity,” *Geopolitics, History, and International Relations*, volume 1, number 2, pp. 41–78.

Nick Bostrom, 2009b. “Dinosaurs, dodos, humans?” *Review of Contemporary Philosophy*, volume 8, 4982, and at https://www.addletonacademicpublishers.com/contents-rcp/112-volume-8-2009/533-dinosaurs-dodos-humans, accessed 10 March 2024.

Nick Bostrom, 2005a. “A history of transhumanist thought,” *Journal of Evolution and Technology*, volume 14, number 1, at http://jetpress.org/volume14/bostrom.html, accessed 10 March 2024.

Nick Bostrom, 2005b. “Letter from Utopia,” *Studies in Ethics, Law, and Technology*, volume 2, number 1, pp. 1–7.
doi: https://doi.org/10.2202/1941-6008.1025, accessed 10 March 2024.

Nick Bostrom, 2005c. “Transhumanist values,” *Journal of Philosophical Research* (supplement), volume 30.
doi: https://doi.org/10.5840/jpr_2005_26, accessed 10 March 2024.

Nick Bostrom, 2003. “Astronomical waste: The opportunity cost of delayed technological development,” *Utilitas*, volume 15, pp. 308–314.
doi: https://doi.org/10.1017/S0953820800004076, accessed 10 March 2024.

Nick Bostrom, 2002. “Existential risks: Analyzing human extinction scenarios and related hazards,” *Journal of Evolution and Technology*, volume 9, at http://jetpress.org/volume9/risks.html, accessed 10 March 2024.

Nick Bostrom and Anders Sandberg, 2009. “Cognitive enhancement: Methods, ethics, regulatory challenges,” *Science and Engineering Ethics*, volume 15, number 3, pp. 311–341.
doi: https://doi.org/10.1007/s11948-009-9142-5, accessed 10 March 2024.

Nick Bostrom and Toby Ord, 2006. “The reversal test: Eliminating status quo bias in applied ethics,” *Ethics*, volume 116, number 4, pp. 656–679.
doi: https://doi.org/10.1086/505233, accessed 10 March 2024.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang, 2023. “Sparks of artificial general intelligence: Early experiments with GPT-4,” *arXiv*:2303.12712 (22 March).
doi: https://doi.org/10.48550/arXiv.2303.12712, accessed 31 January 2024.

Megan Cerullo, 2023. “AI-powered ‘robot’ lawyer won’t argue in court after jail threats,” *CBS* (26 January), at https://www.cbsnews.com/news/robot-lawyer-wont-argue-court-jail-threats-do-not-pay/, accessed 31 January 2024.

Laurie Clarke, 2023. “How Silicon Valley doomers are shaping Rishi Sunak’s AI plans,” *Politico* (14 September), at https://www.politico.eu/article/rishi-sunak-artificial-intelligence-pivot-safety-summit-united-kingdom-silicon-valley-effective-altruism/, accessed 31 January 2024.

Aubrey Clayton, 2020. “How eugenics shaped statistics,” *Nautilus* (27 October), at https://nautil.us/how-eugenics-shaped-statistics-238014/, accessed 31 January 2024.

Devin Coldewey, 2022. “Anthropic’s quest for better, more explainable AI attracts $580M,” *TechCrunch* (29 April), at https://techcrunch.com/2022/04/29/anthropics-quest-for-better-more-explainable-ai-attracts-580m/, accessed 31 January 2024.

Devin Coldewey, 2021. “Anthropic is the new AI research outfit from OpenAI’s Dario Amodei, and it has $124M to burn,” *TechCrunch* (28 May), at https://techcrunch.com/2021/05/28/anthropic-is-the-new-ai-research-outfit-from-openais-dario-amodei-and-it-has-124m-to-burn/, accessed 31 January 2024.

Samantha Cole, 2023. “Largest dataset powering AI images removed after discovery of child sexual abuse material,” *404 Media* (20 December), at https://www.404media.co/laion-datasets-removed-stanford-csam-child-abuse/, accessed 31 January 2024.

Alice Crary, 2023. “The toxic ideology of longtermism,” *Radical Philosophy*, volume 214, pp. 49–57, and at https://www.radicalphilosophy.com/commentary/the-toxic-ideology-of-longtermism, accessed 31 January 2024.

Carla Cremer, 2023. “How effective altruists ignored risk,” *Vox* (30 January), at https://www.vox.com/future-perfect/23569519/effective-altrusim-sam-bankman-fried-will-macaskill-ea-risk-decentralization-philanthropy, accessed 31 January 2024.

Anthony Cuthbertson, 2022. “‘The game is over’: Google’s DeepMind says it is on verge of achieving human-level AI,” *Yahoo Finance* (17 May), at https://finance.yahoo.com/news/game-over-google-deepmind-says-133304961.html, accessed 31 January 2024.

Olivier Dard and Alexandre Moatti, 2017. “The history of *transhumanism* (cont.),” *Notes and Queries*, volume 64, number 1, pp. 167–170.
doi: https://doi.org/10.1093/notesj/gjw256, accessed 10 March 2024.

Jacob Davis, 2023. “Longtermists are pushing a new cold war with China,” *Jacobin* (25 May), at https://jacobin.com/2023/05/longtermism-new-cold-war-biden-administration-china-semiconductors-ai-policy, accessed 1 February 2024.

Neşe Devenot, 2023. “TESCREAL hallucinations: Psychedelic and AI hype as inequality engines,” *Journal of Psychedelic Studies*, volume 7, number S1, pp. 22–39.
doi: https://doi.org/10.1556/2054.2023.00292, accessed 10 March 2024.

Mirtha Donastorg, 2023. “Potentially useful, but error-prone: ChatGPT on the Black tech ecosystem,” *The Plug* (27 February), at https://tpinsights.com/potentially-useful-but-error-prone-chatgpt-on-the-black-tech-ecosystem/, accessed 31 January 2024.

Dante D’Orazio, 2014. “Elon Musk believes colonizing Mars will save humanity,” *The Verge* (4 October), at https://www.theverge.com/2014/10/4/6907721/elon-musks-believes-colonizing-mars-will-save-humanity, accessed 31 January 2024.

Maureen Dowd, 2017. “Elon Musk’s billion-dollar crusade to stop the A.I. apocalypse,” *Vanity Fair* (26 March), at https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x, accessed 31 January 2024.

Sheldon Ekland-Olson, 2011. *Who lives, who dies, who decides? Abortion, neonatal care, assisted dying, and capital punishment*. New York: Routledge.
doi: https://doi.org/10.4324/9780203182277, accessed 10 March 2024.

European Union, 2011. "Directive 2011/65/EU of the European Parliament and of the Council of 8 June 2011 on the restriction of the use of certain hazardous substances in electrical and electronic equipment (recast) Text with EEA relevance," at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011L0065, accessed 1 February 2024.

Grant Fergusson, Caitriona Fitzgerald, Chris Frascella, Megan Iorio, Tom McBrien, Calli Schroeder, Ben Winters, and Enid Zhou, 2023. "Generating harms: Generative AI's impact & paths forward," *Electronic Privacy Information Center*, at https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf, accessed 31 January 2024.

Hayden Field, 2023. "Anthropic, the OpenAI rival, is in talks to raise $750 million in funding at an $18.4 billion valuation," *CNBC* (21 December), at https://www.cnbc.com/2023/12/21/openai-rival-anthropic-in-talks-to-raise-750-million-funding-round.html, accessed 31 January 2024.

McKenna Fitzgerald, Aaron Boddy, and Seth D. Baum, 2020. "2020 survey of artificial general intelligence projects for ethics, risk, and policy," *Global Catastrophic Risk Institute, Technical Report*, 20-1, at https://gcrinstitute.org/papers/055_agi-2020.pdf, accessed 31 January 2024.

Richard S. Fogarty and Michael A Osborne, 2010. "Eugenics in France and the colonies," In: Alison Bashford and Philippa Levine (editors). *Oxford handbook of the history of eugenics*. Oxford: Oxford University Press, pp. 332–346.
doi: https://doi.org/10.1093/oxfordhb/9780195373141.013.0020, accessed 10 March 2024.

Fortune Editors, 2023. "Anthropic's CEO says why he quit his job at OpenAI to start a competitor that just received billions from Amazon and Google," *Fortune* (26 September), at https://fortune.com/2023/09/26/anthropic-ceo-interview-quit-open-ai-amazon-investment/, accessed 31 January 2024.

Kathy Gaca, 2003. *The making of fornication: Eros, ethics, and political reform in Greek philosophy and early Christianity*. Berkeley: University of California Press.

David J. Galton, 1998. "Greek theories on eugenics," *Journal of Medical Ethics*, volume 24, pp. 263–267.
doi: https://doi.org/10.1136/jme.24.4.263, accessed 10 March 2024.

Francis Galton, 1869. *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan & Co.

Timnit Gebru, 2023. "Deep Learning Indaba DAY 2," at https://www.youtube.com/watch?v=5cm_FvHmtVI, accessed 31 January 2024.

Timnit Gebru, 2022. "Effective altruism is pushing a dangerous brand of 'AI safety'," *Wired* (30 November), at https://www.wired.com/story/effective-altruism-artificial-intelligence-sam-bankman-fried/, accessed 31 January 2024.

Timnit Gebru and Margaret Mitchell, 2022. "We warned Google that people might believe AI was sentient. Now it's happening," *Washington Post* (17 June), at https://www.washingtonpost.com/opinions/2022/06/17/google-ai-ethics-sentient-lemoine-warning/, accessed 31 January 2024.

Thomas Germain, 2023. "'Magic intelligence in the sky': Sam Altman has a cute new name for the singularity," *Yahoo Finance* (13 November), at https://finance.yahoo.com/news/magic-intelligence-sky-sam-altman-174500884.html, accessed 31 January 2024.

Ben Goertzel, 2015. "Superintelligence: Fears, promises and potentials: Reflections on Bostrom's Superintelligence, Yudkowsky's From AI to Zombies, and Weaver and Veitas's 'Open-Ended Intelligence'," *Journal of Ethics and Emerging Technologies*, volume 25, number 2, pp. 55–87.
doi: https://doi.org/10.55613/jeet.v25i2.48, accessed 10 March 2024.

Ben Goertzel, 2010. *A cosmist manifesto: Practical philosophy for the posthuman age*. Lexington, Kent.: Humanity Plus Press, and at https://goertzel.org/CosmistManifesto_July2010.pdf, accessed 31 January 2024.

Sharon Goldman, 2023. "Sen. Murphy's tweets on ChatGPT spark backlash from former White House AI policy advisor," *VentureBeat* (28 March), at https://venturebeat.com/ai/sen-murphys-tweets-on-chatgpt-spark-backlash-from-former-white-house-ai-policy-advisor/, accessed 31 January 2024.

Linda Gottfredson, 1997. "Why g matters: The complexity of everyday life," *Intelligence*, volume 24, number 1, pp. 79–132.
doi: https://doi.org/10.1016/S0160-2896(97)90014-3, accessed 10 March 2024.

Mary Gray and Siddharth Suri, 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Boston, Mass.: Houghton Mifflin Harcourt.

Hilary Greaves and William MacAskill, 2019. "The case for strong longtermism," *Global Priorities Institute (GPI) Working Paper*, number 7-2019, at https://web.archive.org/web/20210710220451/https://globalprioritiesinstitute.org/wp-content/uploads/Hilary-Greaves-and-William-MacAskill_strong-longtermism.pdf, accessed 31 January 2024.

Tristan Greene, 2022. "Meta takes new AI system offline because Twitter users are mean," *The Next Web* (19 November), at https://thenextweb.com/news/meta-takes-new-ai-system-offline-because-twitter-users-mean, accessed 31 January 2024.

Peter M. Haas, 1992. "Introduction: Epistemic communities and international policy coordination," *International Organization*, volume 46, number 1, pp. 1–35.
doi: https://doi.org/10.1017/S0020818300001442, accessed 31 January 2024.

Asmelash Teka Hadgu, Paul Azunre, and Timnit Gebru, 2023. "Combating harmful hype in natural language processing," *ICLR 2023*, at https://pml4dc.github.io/iclr2023/pdf/PML4DC_ICLR2023_39.pdf, accessed 31 January 2024.

Karen Hao, 2020. "The messy, secretive reality behind OpenAI's bid to save the world," *MIT Technology Review* (17 February), at https://www.technologyreview.com/2020/02/17/844721/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/, accessed 31 January 2024.

Malcolm Harris, 2023. *Palo Alto: A history of California, capitalism, and the world*. New York: Little, Brown.

Alex Heath, 2024. "Mark Zuckerberg's new goal is creating artificial general intelligence," *The Verge* (18 January), at https://www.theverge.com/2024/1/18/24042354/mark-zuckerberg-meta-agi-reorg-interview, accessed 31 January 2024.

Brian Hepburn and Hanne Andersen, 2021. "Scientific method," In: Edward Zalta (editor). *Stanford Encyclopedia of Philosophy* (1 June), at https://plato.stanford.edu/entries/scientific-method/, accessed

1February 2024.

Richard Herrnstein and Charles Murray, 1994. *The Bell Curve: Intelligence and class structure in American life*. New York: Free Press.

David Hill, 2013. "Exclusive interview: Ray Kurzweil discusses his first two months at Google," *Singularity Hub* (19 March), at https://singularityhub.com/2013/03/19/exclusive-interview-ray-kurzweil-discusses-his-first-two-months-at-google/, accessed 31 January 2024.

Julian Huxley, 1957. *New bottles for new wine, essays*. London: Chatto & Windus.

International Electrotechnical Commission, 2020. "Testing to ensure electronic devices are safe to use in extreme weather," at https://www.iec.ch/blog/testing-ensure-electronic-devices-are-safe-use-extreme-weather, accessed 1 February 2024.

Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru, 2023. "AI art and its impact on artists," *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 363–374.
doi: https://doi.org/10.1145/3600211.3604681, accessed 10 March 2024.

Liwei Jiang, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jenny Liang, Jesse Dodge, Keisuke Sakaguchi, Maxwell Forbes, Jon Borchardt, Saadia Gabriel, Yulia Tsvetkov, Oren Etzioni, Maarten Sap, Regina Rini, and Yejin Choi, 2022. "Can machines learn morality? The Delphi experiment," *arXiv*:2110.07574v2 (12 July).
doi: https://doi.org/10.48550/arXiv.2110.07574, accessed 31 January 2024.

Yarden Katz, 2020. *Artificial whiteness: Politics and ideology in artificial intelligence*. New York: Columbia University Press.
doi: https://doi.org/10.7312/katz19490, accessed 10 March 2024.

Ron Kaufman, 1992. "U. Delaware reaches accord on race studies," *The Scientist* (5 July), at https://archive.ph/XobFu#selection-86.0-98.0, accessed 31 January 2024.

Mehtab Khan and Alex Hanna, 2022. "The subjects and stages of AI dataset development: A framework for dataset accountability," *SSRN* (13 September).
doi: https://dx.doi.org/10.2139/ssrn.4217148, accessed 31 January 2024.

Raffi Khatchadourian, 2015. "The doomsday invention," *New Yorker* (23 November), at https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom, accessed 31 January 2024.

Heidy Khlaaf, 2023. "Towards comprehensive risk assessments and assurance of AI-based systems," *Trail of Bits* (7 March), at https://www.trailofbits.com/documents/Toward_comprehensive_risk_assessments.pdf, accessed 31 January 2024.

Tom Koch, 2020. "Transhumanism, moral perfection, and those 76 trombones," *Journal of Medicine & Philosophy*, volume 45, number 2, pp. 179–192.
doi: https://dx.doi.org/10.1093/jmp/jhz040, accessed 31 January 2024.

Alexander Kossiakoff, Steven Biemer, Samuel Seymour, and David Flanigan, 2020. *Systems engineering principles and practice*. Third edition. New York: Wiley.
doi: https://dx.doi.org/10.1002/9781119516699, accessed 10 March 2024.

Ray Kurzweil, 2006. "Reinventing humanity: The future of human-machine intelligence," *Futurist*, pp. 39–40, 42–46, and at https://www.singularity.com/KurzweilFuturist.pdf, accessed 31 January 2024.

Ray Kurzweil, 2005. *The singularity is near: When humans transcend biology*. New York: Viking.

Ray Kurzweil, 1999. *The age of spiritual machines: When computers exceed human intelligence*. New York: Viking.

Shane Legg, 2008. "Machine super intelligence," Ph.D. dissertation, University of Lugano, at https://www.vetta.org/documents/Machine_Super_Intelligence.pdf, accessed 31 January 2024.

Shane Legg and Marcus Hutter, 2007. "Universal intelligence: A definition of machine intelligence," *Minds and Machines*, volume 17, number 4, pp. 391–444.
doi: https://dx.doi.org/10.1007/s11023-007-9079-x, accessed 10 March 2024.

Gideon Lewis-Kraus, 2022. "The reluctant prophet of effective altruism," *New Yorker* (8 August), at https://www.newyorker.com/magazine/2022/08/15/the-reluctant-prophet-of-effective-altruism, accessed 31 January 2024.

Connie Loizos, 2019. "Sam Altman's leap of faith," *TechCrunch* (18 May), at https://techcrunch.com/2019/05/18/sam-altmans-leap-of-faith/, accessed 31 January 2024.

Alexandra Sasha Luccioni, Yacine Jernite, and Emma Strubell, 2023. "Power hungry processing: Watts driving the cost of AI deployment?" *arXiv*:2311.16863 (23 November).
doi: https://doi.org/10.48550/arXiv.2311.16863, accessed 2 February 2024.

William MacAskill, 2022. *What we owe the future*. New York: Basic Books.

William MacAskill, 2015. *Doing good better: Effective altruism and how you can make a difference*. New York: Gotham Books.

William MacAskill, 2013. "Replaceability, career choice, and making a difference," *Ethical Theory and Moral Practice*, volume 17, pp. 269–283.
doi: https://dx.doi.org/10.1007/s10677-013-9433-4, accessed 10 March 2024.

John Markoff, 2005. "Behind artificial intelligence, a squadron of bright real people," *New York Times* (14 October), at https://www.nytimes.com/2005/10/14/technology/behind-artificial-intelligence-a-squadron-of-bright-real-people.html, accessed 31 January 2024.

Daphne Martschenko, 2017. "The IQ test wars: Why screening for intelligence is still so controversial," *The Conversation* (10 October), at https://theconversation.com/the-iq-test-wars-why-screening-for-intelligence-is-still-so-controversial-81428, accessed 31 January 2024.

Dylan Matthews, 2022. "How effective altruism went from a niche movement to a billion-dollar force," *Vox* (8 August), at https://www.vox.com/future-perfect/2022/8/8/23150496/effective-altruism-sam-bankman-fried-dustin-moskovitz- billionaire-philanthropy-crytocurrency, accessed January 31, 2024.

Dylan Matthews, 2015. "I spent a weekend at Google talking with nerds about charity. I came away ... worried," *Vox* (10 August), at https://www.vox.com/2015/8/10/9124145/effective-altruism-global-ai, accessed 31 January 2024.

John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, 1955. "A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955," (31 August), at https://archive.computerhistory.org/resources/access/text/2023/06/102720392-05-01-acc.pdf, accessed 10 March 2024.

Dan McQuillan, 2022. *Resisting AI: An anti-fascist approach to artificial intelligence*. Bristol: Bristol University Press.

doi: https://doi.org/10.2307/j.ctv2rcnp21, accessed 10 March 2024.

Barry Mehler, 1997. "Beyondism: Raymond B. Cattell and the new eugenics," *Genetica*, volume 99, numbers 2–3, pp. 153–163.
doi: https://doi.org/10.1007/BF02259519, accessed 10 March 2024.

Dan Milmo and Agency, 2023. "Two US lawyers fined for submitting fake court citations from ChatGPT," *Guardian* (23 June), at https://www.theguardian.com/technology/2023/jun/23/two-us-lawyers-fined-submitting-fake-court-citations-chatgpt, accessed 31 January 2024.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis, 2015. "Human-level control through deep reinforcement learning," *Nature*, volume 518, number 7540 (26 February), pp. 529–533.
doi: https://doi.org/10.1038/nature14236, accessed 10 March 2024.

Max More, 1998. "The extropian principles, v. 3.0," at http://www.mrob.com/pub/religion/extro_prin.html, accessed 10 March 2024.

Ernest Mwebaze, Timnit Gebru, Andrea Frome, Solomon Nsumba, and Jeremy Tusubira, 2019. "iCassava 2019 fine-grained visual categorization challenge," *arXiv*:1908.02900 (8 August).
doi: https://doi.org/10.48550/arXiv.1908.02900, accessed 31 January 2024.

Jordan Novet, 2015. "Sam Altman, Elon Musk, Peter Thiel, and others commit $1B to nonprofit artificial intelligence research lab OpenAI," *VentureBeat* (11 December), at https://venturebeat.com/business/sam-altman-elon-musk-peter-thiel-and-others-commit-1b-to-nonprofit-artificial-research-lab-openai/, accessed 31 January 2024.

Scott W. O'Leary-Kelly and Robert J. Vokurka, 1998. "The empirical assessment of construct validity," *Journal of Operations Management*, volume 16, number 4, pp. 387–405.
doi: https://doi.org/10.1016/S0272-6963(98)00020-5, accessed 10 March 2024.

Lorena O'Neil, 2023. "These women tried to warn us about AI," *Rolling Stone* (12 August), at https://www.rollingstone.com/culture/culture-features/women-warnings-ai-danger-risk-before-chatgpt-1234804367/, accessed 31 January 2024.

Toby Ord, 2020. *The precipice: Existential risk and the future of humanity*. New York: Hachette Books.

Patrick Parrinder, 1997. "Eugenics and utopia: Sexual selection from Galton to Morris," *Utopian Studies*, volume 8, number 2, pp. 1–12.

David Pearce, 1995. "The hedonistic imperative," at https://www.hedweb.com, accessed 10 March 2024.

Cassio Pennachin and Ben Goertzel, 2007a. "Contemporary approaches to artificial general intelligence," In: Cassio Pennachin and Ben Goertzel (editors). *Artificial general intelligence*. Berlin: Springer, pp. 1–30.
doi: https://doi.org/10.1007/978-3-540-68677-4_1, accessed 10 March 2024.

Cassio Pennachin and Ben Goertzel (editors), 2007b. *Artificial general intelligence*. Berlin: Springer.
doi: https://doi.org/10.1007/978-3-540-68677-4, accessed 10 March 2024.

Billy Perrigo, 2023. "Exclusive: OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic," *Time* (18 January), at https://time.com/6247678/openai-chatgpt-kenya-workers/, accessed 31 January 2024.

Maria Sophia Quine, 2010. "The first-wave eugenic revolution in southern Europe: Science *sans frontières*," In: Alison Bashford and Philippa Levine (editors). *Oxford handbook of the history of eugenics*. Oxford: Oxford University Press, pp. 377–397.
doi: https://doi.org/10.1093/oxfordhb/9780195373141.013.0023, accessed 10 March 2024.

Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes, 2020. "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33–44.
doi: https://doi.org/10.1145/3351095.3372873, accessed 10 March 2024.

Antonio Regalado, 2018. "A startup is pitching a mind-uploading service that is '100 percent fatal'," *MIT Technology Review* (13 March), at https://www.technologyreview.com/2018/03/13/144721/a-startup-is-pitching-a-mind-uploading-service-that-is-100-percent-fatal/, accessed 31 January 2024.

Ed Regis, 1994. "Meet the extropians," *Wired* (1 October), at https://www.wired.com/1994/10/extropians/, accessed 31 January 2024.

Reuters, 2023a. "OpenAI may leave the EU if regulations bite — CEO" (24 May), at https://www.reuters.com/technology/openai-may-leave-eu-if-regulations-bite-ceo-2023-05-24/, accessed 31 January 2024.

Reuters, 2023b. "Elon Musk launches AI firm xAI as he looks to take on OpenAI" (13 July), at https://www.reuters.com/technology/elon-musks-ai-firm-xai-launches-website-2023-07-12/, accessed 31 January 2024.

Aida Roige, 2014. "Intelligence and IQ testing," *Eugenics Archives*, at https://www.eugenicsarchive.ca/encyclopedia?id=535eecb77095aa000000023a, accessed 31 January 2024.

Kevin Roose, 2023. "Inside the white-hot center of AI doomerism," *New York Times* (11 July), at https://www.nytimes.com/2023/07/11/technology/anthropic-ai-claude-chatbot.html, accessed 31 January 2024.

David Rowan, 2015. "DeepMind: Inside Google's super-brain," *Wired* (22 June), at https://www.wired.co.uk/article/deepmind, accessed 31 January 2024.

Douglas Rushkoff, 2022. *Survival of the richest: Escape fantasies of the tech billionaires*. New York: Norton.

Melia Russell and Julia Black, 2023. "He's played chess with Peter Thiel, sparred with Elon Musk and once, supposedly, stopped a plane crash: Inside Sam Altman's world, where truth is stranger than fiction," *Business Insider* (27 April), at https://www.businessinsider.com/sam-altman-openai-chatgpt-worldcoin-helion-future-tech-2023-4, accessed 31 January 2024.

Stuart Russell and Peter Norvig, 2010. *Artificial intelligence: A modern approach*. Third edition. London: Pearson Education.

Angela Saini, 2019. *Superior: The return of race science*. Boston, Mass.: Beacon Press.

Paula Sambo, Jeremy Hill, and Yueqi Yang, 2023. "FTX halts sale of its $500 million stake in AI startup Anthropic," *Bloomberg* (27 June), at https://www.bloomberg.com/news/articles/2023-06-27/bankman-fried-s-ftx-said-to-halt-sale-of-its-stake-in-ai-startup-anthropic?embedded-checkout=true, accessed 31 January 2024.

Anders Sandberg and Nick Bostrom, 2008. "Whole brain emulation: A roadmap," *Future of Humanity*

Institute, Oxford University, Technical Report, number 2008-3, at http://www.fhi.ox.ac.uk/reports/2008-3.pdf, accessed 31 January 2024.

Chirag Shah and Emily M. Bender 2024. "Envisioning information access systems: What makes for good tools and a healthy Web?" *ACM Transactions on the Web*, version at https://faculty.washington.edu/ebender/papers/Envisioning_IAS_preprint.pdf, accessed 31 January 2024.
doi: https://doi.org/10.1145/3649468, accessed 10 March 2024.

Chirag Shah and Emily M. Bender, 2022. "Situating search," *CHIIR '22: Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, pp. 221–232.
doi: https://doi.org/10.1145/3498366.3505816, accessed 10 March 2024.

Sam Shead, 2020. "How DeepMind boss Demis Hassabis used chess to get billionaire Peter Thiel to 'take notice' of his AI lab," *CNBC* (7 December), at https://www.cnbc.com/2020/12/07/deepminds-demis-hassabis-used-chess-to-get-peter-thiels-attention.html, accessed 31 January 2024.

Sam Shead, 2017. "How DeepMind convinced billionaire Peter Thiel to invest without moving the company to Silicon Valley," *Business Insider* (18 Juiy), at https://www.businessinsider.com/how-deepmind-convinced-peter-thiel-to-invest-outside-silicon-valley-2017-7, accessed 31 January 2024.

Henry Shevlin, Karina Vold, Matthew Crosby, and Marta Halina, 2019. "The limits of machine intelligence: Despite progress in machine intelligence, artificial general intelligence is still a major challenge," *EMBO Reports*, volume 20, number 10, e49177.
doi: https://doi.org/10.15252/embr.201949177, accessed 10 March 2024.

Rachel Shin, 2023. "Elon Musk wants to create a superintelligent A.I. because he thinks a smarter A.I. is less likely to wipe out humanity," *Fortune* (17 July), at https://fortune.com/2023/07/17/elon-musk-superintelligent-a-i-less-likely-to-wipe-out-humanity-chatgpt-openai/, accessed 31 January 2024.

MacKenzie Sigalos, 2023. "Sam Bankman-Fried found guilty on all seven criminal fraud counts," *CNBC* (2 November), at https://www.cnbc.com/2023/11/02/sam-bankman-fried-found-guilty-on-all-seven-criminal-fraud-counts.html, accessed 31 January 2024.

David Silver, Satinder Singh, Doina Precup, and Richard S. Sutton, 2021. "Reward is enough," *Artificial Intelligence*, volume 299, 103535.
doi: https://doi.org/10.1016/j.artint.2021.103535, accessed 10 March 2024.

Robert Sparrow, 2011. "A not-so-new eugenics: Harris and Savulescu on human enhancement," *Hastings Center Report*, volume 41, number 1, pp. 32–42.
doi: https://doi.org/10.1002/j.1552-146x.2011.tb00098.x, accessed 10 March 2024.

Alexandra Minna Stern, 2005. "Sterilized in the name of public health: Race, immigration, and reproductive control in modern California," *American Journal of Public Health*, volume 95, number 7, pp. 1,128–1,138.
doi: https://doi.org/10.2105/AJPH.2004.041608, accessed 10 March 2024.

Alexandra Minna Stern, Nicole L. Novak, Natalie Lira, Kate O'Connor, Siobán Harlow, and Sharon Kardia, 2017. "California's sterilization survivors: An estimate and call for redress," *American Journal of Public Health*, volume 107, number 1, pp. 50–54.
doi: https://doi.org/10.2105/AJPH.2016.303489, accessed 10 March 2024.

Matthew Sweet, 2011. "Introduction," *Francis Galton's Kantsaywhere, UCL Special Collections*, at https://www.ucl.ac.uk/library/special-collections/kantsaywhere, accessed 31 January 2024.

Gillian Tan, Edward Ludlow, Shirin Ghaffary, and Bloomberg, 2023. "Sam Altman's OpenAI to be second-most valuable U.S. startup behind Elon Musk's SpaceX based on early-talks funding round," *Fortune* (23 December), at https://fortune.com/2023/12/23/openai-valuation-100-billion-funding-round/, accessed 31 January 2024.

Max Tegmark, 2017. "Effective altruism, existential risk, & existential hope," at https://www.youtube.com/watch?v=2f1lmNqbgrk, accessed 31 January 2024.

James Temperton, 2017. "DeepMind's new AI ethics unit is the company's next big move," *Wired* (4 October), at https://www.wired.co.uk/article/deepmind-ethics-and-society-artificial-intelligence, accessed 31 January 2024.

David Thiel, 2023. "Identifying and eliminating CSAM in generative ML training data and models," *Stanford Internet Observatory Cyber Policy Center* (20 December), at https://stacks.stanford.edu/file/druid:kh752sm9123/ml_training_data_csam_report-2023-12-20.pdf, accessed 31 January 2024.

Nitasha Tiku, 2023. "How elite schools like Stanford became fixated on the AI apocalypse," *Washington Post* (5 July), at https://www.washingtonpost.com/technology/2023/07/05/ai-apocalypse-college-students, accessed 31 January 2024.

Émile P. Torres, 2024. "Colonization, consciousness, and longtermism," In: Marcello Di Paola and Mirko Garasic (editors). *Philosophy of outer space: Explorations, controversies, speculations*. London: Routledge. doi: https://doi.org/10.4324/9781003374381, accessed 10 March 2024.

Émile P. Torres, 2023a. "Nick Bostrom, longtermism, and the eternal return of eugenics," *Truthdig* (23 January), at https://www.truthdig.com/articles/nick-bostrom-longtermism-and-the-eternal-return-of-eugenics-2/, accessed 31 January 2024.

Émile P. Torres, 2023b. "The acronym behind our wildest AI dreams and nightmares," *Truthdig* (15 June), at https://www.truthdig.com/articles/the-acronym-behind-our-wildest-ai-dreams-and-nightmares/, accessed 31 January 2024.

Émile P. Torres, 2023c. "'Effective accelerationism' and the pursuit of cosmic utopia," *Truthdig* (14 December), at https://www.truthdig.com/articles/effective-accelerationism-and-the-pursuit-of-cosmic-utopia/, accessed 31 January 2024.

Émile P. Torres, 2022. "What 'longtermism' gets wrong about climate change," *Bulletin of the Atomic Scientists* (22 November), at https://thebulletin.org/2022/11/what-longtermism-gets-wrong-about-climate-change/#post-heading, accessed 31 January 2024.

Émile P. Torres, 2021. "The dangerous ideas of longtermism and existential risk," *Current Affairs* (28 July), at https://www.currentaffairs.org/2021/07/the-dangerous-ideas-of-longtermism-and-existential-risk, accessed 31 January 2024.

Émile P. Torres, 2019. "Facing disaster: The great challenges framework," *Foresight*, volume 21, number 1, pp. 4–34.
doi: https://doi.org/10.1108/FS-04-2018-0040, accessed 10 March 2024.

Priyadarshi Tripathy and Kshirasagar Naik, 2011. *Software testing and quality assurance: Theory and practice*. New York: Wiley.

Emily Tucker, 2022. "Artifice and intelligence," *Tech Policy Press* (16 March), at https://www.techpolicy.press/artifice-and-intelligence/, accessed 31 January 2024.

Marius Turda, 2010. "Race, science, and eugenics in the twentieth century," In: Alison Bashford and

Philippa Levine (editors). *Oxford handbook of the history of eugenics*. Oxford: Oxford University Press, pp. 62–79.
doi: https://doi.org/10.1093/oxfordhb/9780195373141.013.0004, accessed 10 March 2024.

Iris van Rooij, Olivia Guest, Federico Adolfi, Ronald de Haan, Antonina Kolokolova, and Patricia Rich, 2023. "Reclaiming AI as a theoretical tool for cognitive science," *PsyArXiv* (1 August).
doi: https://doi.org/10.31234/osf.io/4cbuv, accessed 31 January 2024.

Ashlee Vance, 2016. *Elon Musk: How the billionaire CEO of SpaceX and Tesla is shaping our future*. London: Virgin.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, 2017. "Attention is all you need," *NeurIPS* at https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html, accessed 31 January 2024.

Peter Voss, 2007. "Essentials of general intelligence: The direct path to artificial general intelligence," In: Cassio Pennachin and Ben Goertzel (editors). *Artificial general intelligence*. Berlin: Springer, pp. 131–157.
doi: https://doi.org/10.1007/978-3-540-68677-4_4, accessed 10 March 2024.

Vivek Wadhwa, 2020. "The genetic engineering genie is out of the bottle," *Foreign Policy* (11 September), at https://foreignpolicy.com/2020/09/11/crispr-pandemic-gene-editing-virus/, accessed 31 January 2024.

Elizabeth Weil, 2023. "Sam Altman is the Oppenheimer of our age OpenAI's CEO thinks he knows our future. What do we know about him?" *New York Magazine* (25 September), at https://nymag.com/intelligencer/article/sam-altman-artificial-intelligence-openai-profile.html, accessed 31 January 2024.

H. G. Wells, 1902. *The discovery of the future: A discourse delivered to the Royal Institution on January 24, 1902*. London: T. Fisher Unwin.

H. G. Wells, Julian Huxley, and G.P. Wells, 1931. *The science of life*. London: Cassell.

Sarah Myers West, 2020. "AI and the far right: A history we can't ignore," *AI Now* (4 March), at https://ainowinstitute.org/publication/ai-and-the-far-right-a-history-we-cant-ignore-2, accessed 31 January 2024.

Adrienne Williams, Milagros Miceli, and Timnit Gebru, 2022. "The exploited labor behind artificial intelligence," *Noema* (13 October), at https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/, accessed 31 January 2024.

Chloe Xiang, 2023. "'He would still be here': Man dies by suicide after talking with AI chatbot, widow says," *Vice* (30 March), at https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says, accessed 31 January 2024.

George M. Young, 2012. *The Russian cosmists: The esoteric futurism of Nikolai Fedorov and his followers*. Oxford: Oxford University Press.
doi: https://doi.org/10.1093/acprof:oso/9780199892945.001.0001, accessed 10 March 2024.

Eliezer Yudkowsky, 2023. "Pausing AI developments isn't enough. We need to shut it all down," *Time* (29 March), at https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/, accessed 31 January 2024.

Eliezer S. Yudkowsky, 2000. "The singularitarian principles, version 1.0,," at https://web.archive.org/web/20000621223020/http://singinst.org/singularitarian/principles.html, accessed

10 March 2024.

Tukufu Zuberi, 2001. *Thicker than blood: How racial statistics lie*. Minneapolis: University of Minnesota Press.

Ethan Zuckerman, 2024. “Two warring visions of AI,” *Prospect* (16 January), at https://www.prospectmagazine.co.uk/ideas/technology/64491/two-warring-visions-of-artificial-intelligence-tescreal, accessed 31 January 2024.

Editorial history

Received 21 September 2023; accepted 10 March 2024.

The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence
by Timnit Gebru and Émile P. Torres.
First Monday, volume 29, number 4 (April 2024).
doi: https://dx.doi.org/10.5210/fm.v29i4.13636.