# TESCREAL
## Dr. Émile P. Torres
### Penultimate Draft; don't cite without permission

## Summary

DeepMind, OpenAI, Anthropic, and other leading AI companies are racing to build artificial general intelligence, or AGI. What is motivating this race? Who founded these companies, and why? In a 2024 paper, Timnit Gebru and Émile P. Torres argue that the AGI race emerged out of the TESCREAL movement. This movement dates back to the early 1990s, although it has roots in 20th-century eugenics and 19th-century Russian Cosmism. Over the past decade or so, TESCREALism has become enormously influential within Silicon Valley, embraced and promoted by some of the most powerful figures in the tech world. At its core is a techno-utopian philosophy according to which advanced technologies will enable us to create a new species of "posthumans" who spread beyond Earth to colonize the accessible universe. AGI is integral to realizing this grand eschatological vision: once AGI becomes superintelligent, humanity could delegate it the task of "paradise-engineering," to quote a leading figure of the movement, Nick Bostrom. Without AGI, utopia will likely be impossible, and hence we must build AGI as quickly as we can, while ensuring that it can be controlled by those who build it. The AGI race thus emerged to fulfill the cosmic mission of realizing utopia.

There are many ways for critics to approach the topic of AGI. Some have focused on the environmental impact of large language models (LLMs), which power systems like ChatGPT, seen by most TESCREALists as the stepping stones to AGI. Others point to phenomena like intellectual property (IP) theft, worker exploitation, and AI-generated deepfakes and disinformation. However, a growing number of scholars are beginning to examine the underlying techno-utopian ideologies that have inspired, launched, sustained, and accelerated this race. The TESCREAL framework provides a powerful new way of understanding and critiquing the AGI race, focusing not on its societal consequences but on its root causes. Critics of TESCREALism may thus argue that this framework provides an indispensable tool for addressing the harms of AI, and hence that understanding the TESCREAL ideologies is crucial for combating the ongoing rush to build machines far more "intelligent" than all of humanity combined.

**Keywords** TESCREAL, artificial general intelligence, utopia, human extinction

# 1. Introduction

The "TESCREAL" acronym was coined in early 2023 by the philosopher Émile P. Torres in a collaboration initiated by the computer scientist Timnit Gebru. This project culminated in a paper published the following year titled, "The TESCREAL Bundle: Eugenics and the Promise of Utopia through Artificial General Intelligence," which popularized the acronym. It denotes seven ideologies: transhumanism, Extropianism, singularitarianism, Cosmism, Rationalism, Effective Altruism, and longtermism. Gebru and Torres argue that these ideologies are yet another iteration of the "eternal return of eugenics,"[1] and that they have played an integral role in launching, sustaining, and accelerating the race to build artificial general intelligence, or AGI. On why Gebru and Torres introduced the acronym, they write:

> 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.[2]

This gestures at two interpretations of the TESCREAL thesis. A "weak" interpretation would suggest that one cannot give a complete explanation of the (origins of the) race to build AGI without reference to these seven ideologies. A "strong" interpretation adds that one should understand these ideologies as forming a cohesive bundle or family of worldviews built around shared techno-futuristic themes, common epistemic and moral commitments, and genealogical roots in 20th-century Anglo-American eugenics.[3] A primary impetus behind the TESCREAL concept was to outline a framework from and within which to critique the corresponding social, cultural, and intellectual communities.

In what follows, we begin with a brief examination of the TESCREAL ideologies, and then explore some reasons for accepting the strong TESCREAL thesis that Gebru and Torres defend. After this, we will turn to their connections to the AGI race and conclude with a discussion of whether "safe" AGI is possible, and how the TESCREAL ideologies pose a direct threat to not just marginalized communities but humanity itself.

## 2. What Are the TESCREAL Ideologies?

Transhumanism, on one account, is the claim that we should develop advanced technologies to radically reengineer the human organism, resulting in one or more "posthuman" species. The 2003 Transhumanist FAQ defines it as

> the intellectual and cultural movement that affirms the possibility and desirability of fundamentally improving the human condition through applied reason, especially by developing and making widely available

technologies to eliminate aging and to greatly enhance human intellectual, physical, and psychological capacities.[4]

In "Transhumanist Values," Nick Bostrom identifies the "core value" of transhumanism as being "the opportunity to explore the transhuman and posthuman realms."[5] His subsequent "Letter from Utopia" depicts a future in which our posthuman descendants have overcome aging, gained cyberimmortality by uploading their minds to computers, expanded their cognitive capacities far "beyond the bounds of any genius of humankind," and created a world of "surpassing bliss and delight" marked by so much pleasure that "we sprinkle it in our tea."[6] This eschatological vision of techno-utopia through "person engineering" is the nucleus around which the TESCREAL movement revolves, which is why Torres characterizes transhumanism as the "backbone" of the ideological bundle.[7]

Transhumanism is a radical form of eugenics. Coined in 1883 by Frances Galton, the word "eugenics" derives from Greek, meaning "good" (*eu-*) "birth" (*genos*). The goal of eugenics is to improve human populations through various mechanisms, explored below. We begin with a brief overview of the historical development of eugenics, and then turn to how it gave rise to transhumanism.

### *The Eugenic Roots of Transhumanism*

Proto-eugenics, as we can call it, dates back to the origins of the Western tradition. For example, the "Twelve Tables" of Roman law, ratified in 449 BCE, "made provisions for infanticide on the basis of deformity and weakness."[8] In the first century CE, the Roman philosopher Seneca wrote that

> we put down mad dogs; we kill the wild, untamed ox; we use the knife on sick sheep to stop their infecting the flock; we destroy abnormal offspring at birth; children, too, if they are born weak or deformed, we drown. Yet this is not the work of anger, but of reason—to separate the sound from the worthless.[9]

In books V and VI of the 4th-century-BCE dialogue *Republic*, Plato proposed a utopian society in which "women and men would be matched by the state according to their qualities, like sporting dogs or horses bred for their strength or speed. … Inferior children would be relegated to a working class and discouraged from breeding."[10] Those children identified as "visibly defective" would "be secretly taken away by officials and almost certainly left to die."[11] Aristotle presented his own eugenic roadmap, arguing that human excellence is inherited, and that lawgivers should arrange marriages, control when couples have children, and "make sure that the bodies of the newborns are as they wish them to be." This advice aimed to ensure "citizens and their wives give birth to children who are preferably male and have the same kind of natural character traits that make moral development most likely to succeed."[12] Similar ideas can be found during the Middle Ages and Enlightenment, as when the 18th-century French historian and jurist Guillaume Poncet de la Grave discouraged miscegenation, arguing that "it exposed French blood to corruption and produced disfigured children."[13]

Proto-eugenics transitioned to what scholars call "first-wave eugenics" following the 1869 publication of Francis Galton's *Hereditary Genius*.[14] This marked the first "scientific" treatment of eugenics, as Galton built his eugenic proposals on the theory of evolution by natural selection delineated in Charles Darwin's *On the Origin of Species*, published 10 years earlier. Darwin, a second cousin of Galton's, motivated the idea of natural selection by pointing to our use of "artificial selection," i.e., selective breeding, to modify the phenotypic traits of domesticated plants and animals over time. Galton contended that similar interventions could shape the human species: given that "natural abilities" are heritable, he claimed, if those with "superior faculties" were to produce more offspring than their peers, the constitution of the "human stock" will improve overall.[15]

To achieve this goal, Galton encouraged superior individuals to produce larger families, a strategy known as "positive" eugenics. This inspired the popular "better baby" and "fitter family" contests held throughout the US during the early 20th century. In contrast, "negative" eugenics, which Galton did not emphasize in his work, strives to prevent "unfit" individuals, who were believed to concentrate among the poor, immigrants, and racially minoritized populations—and who were variously labeled "imbeciles," "defectives," "idiots," "congenital invalids," "morons," and "feeble-minded," often identified through IQ tests[16]—from having children. Apart from enforcing racialized immigration bans and segregation of classes and races, an additional way to achieve this is through forced sterilization.[17]

By altering society-wide patterns of reproduction, eugenicists hoped to enhance the overall health and abilities of the "human stock." Eugenicists, who were "anxiously facing globalization's modern advent and growing independence and abolition struggles around the world," as Anita Say Chan observes, also feared that *dysgenic* reproductive patterns, whereby the "unfit" out-breed their more "fit" peers, could result in evolutionary degeneration, an idea that triggered considerable anxiety around the turn of the 20th century, thanks partly to popular works like H. G. Wells' *The Time Machine* (1895).[18]

As Alison Bashford and Philippa Levine observe, "eugenic thought and practice swept the world from the late nineteenth to the mid-twentieth century in a remarkable transnational phenomenon."[19] People on both sides of the political spectrum embraced eugenic thinking, including social progressives, socialists, and communists on the left (e.g., J. B. S. Haldane) and far-right fascists in Germany (e.g., the Nazis). California's eugenics program, for example, was implemented in 1909 during the Progressive Era, and explicitly served as a template for the program established in 1930s Nazi Germany. Even more remarkably, eugenics continued to be practiced after the atrocities of the Second World War came into clear view. California's program, for instance, was not officially terminated until 1979, and some countries previously occupied by the Nazis initiated their own programs after the war concluded. Japan enacted its Eugenic Protection Law in 1948, which resulted in roughly 25,000 people being sterilized between its enactment and abolition in 1996. There is, to paraphrase some scholars, *continuity* of eugenics across the 20th century.[20]

By the 1970s, eugenics came under increasing scrutiny and criticism, though it had been fiercely contested by scientists at least since the 1930s.[21] However, it was around the same time, in the post-war era, that "second-wave eugenics emerged as an offshoot of genetics and biotechnology."[22] Genetic engineering and biotechnology suggested a new way of modifying the phenotypes of individuals by directly altering one's genes or the genes of one's children. (As discussed below, a growing number of people in the 1980s and 90s

argued that molecular nanotechnology and artificial intelligence (AI) could provide additional means of "enhancing" human beings.) Such technologies could enable us to induce significant changes to our bodies and brains over relatively short periods of time: one's own lifetime or, in the case of "designer babies," a single generation. In contrast, selective breeding requires many generations to induce appreciable phenotypic changes.

The key difference between first- and second-wave eugenics thus concerns their respective methodologies: the former relies on positive and negative strategies to shape humanity on transgenerational timescales, whereas the latter employs advanced technology to take immediate control of our evolutionary trajectory. This corresponds to a common distinction between "liberal" and "authoritarian" eugenics. Advocates of liberal eugenics claim to "uphold the principles of bodily autonomy and procreative liberty," emphasizing "individual freedom and individual choice in the area of enhancement technologies."[23] Some proposed a fundamental right they call "morphological freedom," understood "as an extension of one's right to one's body, not just self-ownership but also the right to modify oneself according to one's desires."[24]

These ideas contrast with the "authoritarian" approach of first-wave eugenics, especially in its negative form, which involved top-down interventions from the state to force changes in reproductive patterns across society. Liberal eugenicists thus contend that their form of eugenics purges itself of the features that made authoritarian eugenics morally repugnant. It advocates freedom and choice whereas the latter often utilizes force and coercion. It also rejects the discriminatory attitudes that animated first-wave eugenicists. As Bostrom writes, "racism, sexism, speciesism, belligerent nationalism, and religious intolerance are unacceptable."[25] Ironically, these words were published just 7 years after he declared in an email to the Extropians mailing list (see below) that "Blacks are more stupid than whites. I like that sentence and think it is true," after which he wrote the N-word.[26] We will see that prejudices like those Bostrom lists are, in fact, pervasive within the TESCREAL movement.

### *Early Versus Modern Transhumanism*

Transhumanism differs most saliently from traditional eugenics with respect to its ultimate goal. Whereas eugenicists of both the first and second waves aimed to improve the overall health, intelligence, etc. of our species and/or prevent evolutionary degeneration, transhumanists take this a step further in advocating for the creation of a new population of "superior" beings, often called "posthumans" by contemporary transhumanists. In other words, traditional eugenicists strove to create the very best version of our species possible; transhumanists strive to create an entirely new species.

The history of transhumanism can be divided into two phases: "early" transhumanism was developed by leading eugenicists such as Haldane, J. D. Bernal, and Julian Huxley. An important contribution was Huxley's 1927 book *Religion Without Revelation*, in which he pointed out that we systematically reorganize our environment to suit our desires; however, this "outlook" of bending the world to our has become extendable to humanity itself. "The study of heredity and population-growth," he writes, "and the knowledge of eugenics and of birth-control are pointing the way to wholly new aims—to a conscious control by man of his own nature and racial destiny."[27] In *New Bottles for New Wine*, published 30 years later, Huxley declared that

the human species can, if it wishes, transcend itself—not just sporadically, an individual here in one way, an individual there in another way, but in its entirety, as humanity. We need a name for this new belief. Perhaps transhumanism will serve: man remaining man, but transcending himself, by realizing new possibilities of and for his human nature. ... "I believe in transhumanism": once there are enough people who can truly say that, 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.[28]

The hallmark of early transhumanism is that it combined the *goal* of "transcendence" with the *methodology* of first-wave eugenics. Through positive and negative strategies (Huxley himself favored both[29]), we may eventually produce a population of beings that are in some sense still "man" yet instantiate a fundamentally "new kind of existence." (Peking Man is a variant of *Homo erectus*, and hence Huxley must be suggesting the creation of a new species.)

In contrast, "modern" transhumanism integrates this lofty goal with the novel methodology of *second-wave* eugenics. It prescribes the creation of a new posthuman species through the application of "radical human enhancement" technologies, which could take the form of genetic engineering, biotechnology, synthetic biology, molecular nanotechnology, and AI—which Ray Kurzweil later bundled under the acronym "GNR," standing for "genetics, nanotech, and robotics."[30] For example, brain-computer interfaces could enable us to link our cognitive systems to the Internet; life-extension technologies could give us indefinitely long lives; and molecular nanotechnology could make it possible to upload our minds to silicon computer hardware. This is the technological path to utopia, whereby we transcend our "biological limitations" to attain endless pleasures, immortality, and radically augmented cognitive systems.[31] Some transhumanists refer to the goal of radical enhancement as "The Three Supers," namely, superwellbeing, superlongevity, and superintelligence.[32]

### *Extropianism*

The first organized movement built around modern transhumanist ideology was Extropianism. This emerged from the writings of Max More (born Max O'Connor), who cofounded the Extropy Institute with T. O. Morrow (born Tom Bell) in 1991. The term "extropy" was intended to be a metaphorical antonym of "entropy," defined by More as "the process of increasing intelligence, information, usable energy, life, experience, and growth."[33] Politically, the Extropians were ardent libertarians. Ayn Rand's *Atlas Shrugged* was included on its official reading list, and "barely an issue of [their publication] *Extropy* was published without reference to Ayn Rand." As Alexander Thomas notes, rightwing economists like "Friedrich Hayek, Milton Friedman and Ludwig von Mises were regularly approvingly cited in Extropian circles and radical ideas 'such as privatising the air and the oceans, were often discussed and endorsed by extropians in the 90s.'"[34]

A foundational essay outlining the core tenets of Extropianism was published by More in 1990. It included principles like: (1) Boundless Expansion, which emphasized the acquisition of "more intelligence ... and personal power, an unlimited lifespan, and [the]

removal of [all] limits to self-actualization." (2) Self-Transformation, "both moral and cognitive," which advocated for "biological and neurological augmentation" and a "rejection of central control and maximum sustainable freedom." (3) Dynamic Optimism, which promotes "a positive, empowering attitude towards our individual future and that of all intelligent beings." And (4) Intelligent Technology, which affirmed "the role of science and its offspring, technology, guided by extropian values, in realizing the optimistic, dynamic value-perspective of extropianism."[35] Two years later, More added "Spontaneous Order" to his list, which "explicitly affirmed that self-regulating orderly systems like the free market should be embraced, as they are more likely to intelligently engender extropian goals than human regularity bodies."[36] The first two letters of these five principles spell out the imperative: "BEST DO IT SO!"

In 2004, More published Version 1.0 of "The Proactionary Principle" on his personal website. His aim was to provide an alternative to the Precautionary Principle, which he saw as "strongly biased against the technological progress so vital to the continued survival and well-being of humanity." More's alternative principle affirms that "people's freedom to innovate technologically is valuable to humanity," and hence that "the burden of proof … belongs to those who propose restrictive measures." He warns that cognitive biases may distort our evaluation of technological risks, and argues that "if the precautionary principle had been widely applied in the past, technological and cultural progress would have ground to a halt," resulting in life having "remained poor, nasty, brutish, and short."[37] In later iterations, More foregrounds the idea of "perpetual progress," stating that "progress should not bow to fear but should proceed with eyes wide open."[38]

There are few who would call themselves "Extropians" today. However, the legacy of this ideology within the TESCREAL movement has been significant. Historically, Extropianism brought together transhumanists from around the world through its mailing list, described as "the longest running transhumanist email list in the world."[39] This enabled people like Nick Bostrom, Ben Goertzel, Anders Sandberg, and Eliezer Yudkowsky—names that we will encounter again below—to exchange ideas about the Singularity, human enhancement, and other techno-futuristic topics.[40] The Extropians also organized events like the Vital Progress Summits and EXTRO conferences, which included Ray Kurzweil, Marvin Minsky, and Hans Moravec as speakers.[41] If not for Extropianism, modern transhumanist thought might not have consolidated into a powerful movement within Silicon Valley.

Furthermore, key ideas and themes from Extropianism have shaped other TESCREAL ideologies. The Proactionary Principle and notion of Dynamic Optimism are perhaps best exemplified by "effective accelerationism" (e/acc), a prominent school of thought within the TESCREAL movement (see below), while the emphasis on eradicating cognitive biases and improving our rationality is the central goal of Rationalists. The TESCREAL movement is also largely committed to a libertarian politics, though some AI "doomers" make an exception for government regulation with respect to "frontier" AI models. As Thomas notes, "transhumanists are broadly split between two poles: the right-leaning techno-libertarian wing, often associated with Silicon Valley, and the left-leaning techno-progressive faction most notably represented by transhumanist James Hughes."[42] Although both descended from eugenics, it is the former that Gebru and Torres aim to highlight with their acronym, as the techno-libertarian wing that directly emerged out of Extropianism has become far more influential, including within the ongoing AGI race, than

the techno-progressive camp of Hughes.[43] Extropianism was included in the acronym, therefore, because of its sizable intellectual inheritance, passed down to subsequent ideologies in the bundle. To say that transhumanism is the "backbone" of the TESCREAL movement is thus to say that Extropianism—libertarian modern transhumanism—is at the core of this entire futuristic worldview.

### *Singularitarianism*

The word "singularitarian" was coined in 1991 by the Extropian T. O. Morrow. A singularitarian is one who "believes that the Singularity is possible, that the Singularity is a good thing, and that we should help make it happen."[44] The Singularity is a hypothetical future event with several possible properties: first, it could mark a new epoch in cosmic history whereby the rate of technological change occurs so quickly that humans are unable to comprehend or make sense of the phantasmagoria around them. Imagine 500 years of change happening every minute. Second, it could refer to an "intelligence explosion," whereby AI begins to recursively self-improve, resulting in an artificial superintelligence (ASI) that emerges "within minutes, hours or days."[45] Or third, it could denote a "future that is weirder by far than most science fiction, a difference-in-kind that goes beyond amazing shiny gadgets" due to radical improvements in "human intelligence." On this view, associated with the science fiction writer Vernor Vinge, the future is hidden behind an "event horizon."[46] These are mutually compatible, though many TESCREALists seem to focus primarily on the second interpretation, with the first and third being byproducts of the advent of ASI.

Like Extropians, singularitarians tend to embrace an optimistic view of the future, though they also caution against "existential risks" associated with advanced technologies.[47] In the 1990s and early aughts, Eliezer Yudkowsky was a leading singularitarian, founding the Singularity Institute for Artificial Intelligence (later rebranded as the Machine Intelligence Research Institute) in 2000 with more than $1.6 million from billionaire Peter Thiel.[48] The mission of this institute was "to accelerate toward artificial intelligence."[49] Yudkowsky subsequently became a prominent doomer, arguing with >95% certainty that building AGI—which could quickly self-improve to become ASI—in the near future would result in human annihilation.[50]

Another prominent singularitarian is Ray Kurzweil, who presented at and sponsored the EXTRO 5 conference (2001), delivered a keynote address at the Vital Progress Summit (2004), and "served on the Council of Advisors of the Extropy Institute."[51] In 2005, he published *The Singularity Is Near*, which argued that humans will merge with machines during the Singularity, after which the posthumans we become or create will spread beyond Earth and colonize the universe. Eventually, the universe itself "wakes up" as the light of consciousness floods our future light cone. He describes this futurology as "a new religion," and argues that if you choose not to become posthuman, then "you won't be around for very long to influence the debate."[52] In other words, our species will soon be replaced by our posthuman successors, so the only way to continue the conversation is to embrace radical human enhancements (see subsection 5.4). Based on extrapolations of exponential technological growth, Kurzweil predicts the Singularity will happen in 2045, though other singularitarians have proposed different dates. Yudkowsky, for example, once prognosticated that it will occur in 2025.[53]

Kurzweil's ideas have had a significant impact on the development of TESCREALism, and the singularitarian ideology has influenced many founders of leading AI companies, including Shane Legg,[54] Sam Altman,[55] Dario Amodei,[56] and Elon Musk, the last of whom posted on X, his social media website, in early 2025 that "we are on the event horizon of the singularity" (perhaps, then, Yudkowsky was right).[57]

### *Cosmism*

The provenance of Cosmism lies in the writings of late 19th- and early 20th-century Russian theorists like Nikolai Fedorov (sometimes spelled Fyodorov) and Konstantin Tsiolkovsky. In Fedorov's posthumously published "Philosophy of the Common Task" (1906), he adumbrated a project to create "the technological, social, and political conditions under which it would be possible to resurrect by technological and artificial means all people who have ever lived."[58] Subsequent thinkers combined this vision with revolutionary anarchist and Marxist ideas to produce a variant called "Biocosmism," which stripped the religious elements of Fedorov's project "while still advancing its technoscientific vision of immortality, resurrection for all, and freedom of travel in universal space."[59] As Thomas notes, Fedorov "is considered by some contemporary Russian immortalists to be the first transhumanist."[60]

Russian Cosmism and Biocosmism contrasts with what could be called "modern" Cosmism, which Gebru and Torres intend to highlight with their "TESCREAL" acronym. The most prominent exponent of modern Cosmism is the computer scientist Ben Goertzel, an Extropian transhumanist who writes about the Singularity and founded SingularityNET to advance the goal of building "a decentralized, democratic, inclusive and beneficial Artificial General Intelligence."[61] When asked about whether super-wealthy tech leaders will act benevolently once AGI arrives, Goertzel replied: "Once AGI has obsoleted money the trillionaire overlords will pretty much be frolicking in post-Singularity utopia along with all the rest of us."[62] With respect to the term "Cosmism," he says that "previous users of the term … held views quite sympathetic to my own, so classifying my own perspective as an early 21st century species of Cosmism seems perfectly appropriate."[63]

In 2010, Goertzel published *A Cosmist Manifesto*, which outlined a normative futurology that goes "far beyond" the parochial focus of transhumanism and Extropianism. It prescribes not just radically reengineering humanity, but spreading beyond Earth to redesign the universe itself. Goertzel discusses brain-computer interfaces, virtual reality worlds, sentient AI systems, and "cyberimmortality" through mind-uploading. He also lists "Ten Cosmist Convictions," which elaborate principles outlined in the "Order of Cosmic Engineers," written by fellow Cosmist Giulio Prisco. These declare that "humans will merge with technology, to a rapidly increasing extent," a central theme of Kurzweilian singularitarianism. Those who upload their minds will become immortal, and our posthuman descendants will "spread to the stars and roam the universe." We could even "roam to other dimensions of existence as well, beyond the ones of which we're currently aware," and design "synthetic realities" (virtual worlds) in which sentient beings will live. Scientific "future magic" will enable us to engineer spacetime itself, thus enabling us to achieve, "by scientific means, most of the promises of religions," including resurrecting "the dead by 'copying them into the future.'" This paradisiacal future world will yield "abundances of wealth," and "all of these changes will [lead] to states of individual and

shared awareness possessing depth, breadth and wonder far beyond that accessible to 'legacy humans.'"[64]

Integral to this utopian project is the creation of what Goertzel variously calls "artificial general intelligence" (AGI) or "transhuman minds." He suggests that we have a moral obligation to bring AGI into existence, but also accepts that "there are risks in creating superhuman minds." However, echoing his Extropian colleagues, he declares that "Cosmism is not about faint-heartedly fearing growth because it comes with risks. … Transhuman AGI? Bring it on!"[65]

Gebru and Torres include Cosmism in their acronym precisely because of this connection. Recall that their interest is in the ideologies driving the AGI race. Goertzel played an important early role in promoting the idea of AGI and was "heavily involved in the formation and growth" of the field of AGI R&D.[66] His interest in AGI appears to directly arise from his transhumanist, Extropian, and Cosmist convictions. In fact, Goertzel's work is the main reason that "AGI" has become an entry in our shared lexicon of AI terminology: during the mid-2000s, he was preparing an edited volume with Cassio Pennachin on advanced AI. This included contributions from people like Yudkowsky, the blogger who Goertzel later worked for as the Director of Research at Yudkowsky's Singularity Institute.[67] The tentative title of the book was *Real AI*, but Goertzel wanted something catchier. Soliciting suggestions, a former employee of his, Shane Legg, suggested "artificial general intelligence." Goertzel liked this and changed the title to *Artificial General Intelligence* (2005/2007), thus popularizing the term.[68] Three years later, Legg cofounded the AI company DeepMind with the explicit goal of building AGI. As discussed more in subsection 4.1, initial funding for DeepMind came from Peter Thiel, who had previously funded Yudkowsky's Singularity Institute and, in 2006, cofounded the Singularity Summit with Yudkowsky and Kurzweil, at which figures like Goertzel, Shane Legg, Max More, Nick Bostrom, Vernor Vinge, James Hughes, Anders Sandberg, Steven Pinker, and Demis Hassabis (another cofounder of DeepMind) gave talks.[69]

### Rationalism and Effective Altruism

Rationalism and Effective Altruism (EA) are closely linked. Both emerged around 2009, emphasize "reason" and "rationality," tend to embrace IQ realism, and are inclined toward totalist utilitarian ethics and decision-theoretic notions like expected value theory. The Rationalist and EA communities overlap considerably, and both were (co)founded by people previously embedded within the modern transhumanist movement: Eliezer Yudkowsky, in the case of Rationalism, and Toby Ord, in the case of EA.

The Rationalist movement grew out of the community blogging website LessWrong, started by Yudkowsky circa 2009. This website describes itself as "an online forum and community dedicated to improving human reasoning and decision-making." It states that "many members … are heavily motivated by trying to improve the world as much as possible," which is also the central thrust of EA, and that they became "convinced many years ago that AI was a very big deal for the future of humanity." Hence, "the LessWrong team … are predominantly motivated by trying to cause powerful AI outcomes to be good."[70]

Contributors to LessWrong include many people mentioned above and below: Max More, Ben Goertzel, Toby Ord, Nick Bostrom, Anders Sandberg, Nick Beckstead, and William

MacAskill. Culturally, LessWrong achieved something similar to the Extropians mailing list in the 1990s by bringing together like-minded futurists, thereby enabling a social community to coalesce around the ideology of Rationalism. The impetus behind this ideology can be reconstructed as follows: "paradise-engineering" (Bostrom's term[71]) will require advanced science and technology; advanced science and technology will require superior rationality; hence, by enhancing our rationality we can improve our odds of engineering paradise. Rationality "training" is thus a primary goal of the LessWrong community. It is also a central theme of many publications by Yudkowsky, including his fan-fiction novel *Harry Potter and the Methods of Rationality* and his 1.8-million-word tome on BDSM, decision theory, and Dungeons and Dragons.[72]

Whereas Rationalists strive to optimize their rationality, Effective Altruists (EAs) strive to optimize their morality.[73] They claim to use science and reason to determine the best ways of "doing good better."[74] This has led to counterintuitive conclusions, such as that one should not donate to disaster relief funds or grassroots organizations,[75] but one should support sweatshops.[76]

Some EAs like MacAskill have defended an idea dubbed "earning to give," whereby one pursues lucrative jobs to acquire more money to donate to charities. If working for what MacAskill calls "immoral organizations," such as petrochemical companies, arms dealers, or Wall Street firms, enables one to give more, then this could be a better option than taking a job at a nonprofit charity.[77]

Until his catastrophic downfall, the great success story of earning to give was Sam Bankman-Fried, whose meeting with MacAskill in 2012 altered the course of his life.[78] After graduating from MIT, Bankman-Fried worked at Jane Street Capital (where other EAs, including his brother Gabe, had also worked). He then took a position at MacAskill's Centre for Effective Altruism, which shared office space at Oxford with value-aligned organizations like Our World in Data and Bostrom's now-defunct Future of Humanity Institute, after which he founded the cryptocurrency exchange platform FTX to "get filthy rich, for charity's sake."[79] Like many EAs, Bankman-Fried was deeply influenced by totalist utilitarianism, which posits that the moral rightness or wrongness of an act depends entirely on its consequences; a variant of this claims that acts are right when they maximize *expected value*.[80] This is why working for "immoral organizations" may be not just permissible but morally obligatory: in expectation, one could—depending on the exact numbers—generate more welfare by working on Wall Street than for a nonprofit. This points to a problem with practical applications of expected value theory: there are often no good bases for assigning values and/or probabilities to possible outcomes. One can, therefore, wiggle the numbers to fit just about any desired conclusion.

Totalist utilitarianism also inclines one to accept a quantitative approach to ethics.[81] Consequently, as Torres notes, EA essentially reduces the domain of morality to a branch of economics, where moral decisions become calculations, people become mere fungible containers for value, and value is quantified into mathematically manipulable units.[82] An illustration of this approach comes from Yudkowsky. He argues that, in a forced-choice situation, we should prefer a single person being tortured for 50 years over an inscrutably large number of people suffering the nearly imperceptible discomfort of having a speck of dust in their eye. If one does the math, the second scenario is worse, and hence we should favor the former. Yudkowsky sloganizes this idea as: "Shut up and multiply."[83]

Another example involves the Repugnant Conclusion, or the idea that a world full of trillions of people with barely worthwhile lives is better than one in which a smaller number of people are extremely happy. If the total amount of welfare—aggregated over all those who exist—is greater in the first world than the second, then we should choose the first. This conclusion follows from the axiological component of totalist utilitarianism—namely, totalism or the "Total View"[84]—but despite its "repugnance," EAs like Ord and MacAskill argue that we should not give it too much weight. "The intuition that the Repugnant Conclusion is repugnant may be unreliable," they write, and hence MacAskill concludes (in a separate publication) that "the fact that a theory of population ethics entails the Repugnant Conclusion shouldn't be a decisive reason to reject that theory."[85] Totalism, we will see, is a central component of longtermism, which emerged out of the EA movement.

### *Longtermism*

The longtermist "ethic" comes in two varieties: "moderate" longtermism, as we can call it, is "the view that positively influencing the longterm future is a key moral priority of our time." "Radical" longtermism (sometimes called "strong" longtermism) replaces the indefinite with the definite article before the word "key." It asserts that "positively influencing the longterm future is *the* key moral priority of our time."[86] Most leading longtermists endorse, or are most sympathetic to, radical longtermism. This is what Gebru and Torres primarily focus on in discussing the TESCREAL bundle. In what follows, we will use "longtermism" to refer to "radical longtermism."

One way to understand longtermism is that it is what results when the imperatives of EA collide with facts about our universe uncovered by modern cosmology. According to cosmologists, the universe is enormous: there are between 100 and 400 billion stars in the Milky Way galaxy, and between 200 billion and 2 trillion galaxies in the universe. Along the temporal rather than spatial dimension, Earth will remain habitable for roughly 1 billion years, but if we spread beyond Earth, digital life could persist for perhaps $10^{100}$ years, when the heat death of the universe is scheduled to happen. This means that the future posthuman population could be astronomically large. Carl Sagan estimates 500 trillion future people on Earth if we survive for the next 10 million years;[87] Toby Newberry calculates $10^{45}$ digital posthumans per century if we colonize the Milky Way;[88] and Nick Bostrom says there could be at least $10^{58}$ digital posthumans within our future light cone.[89]

If one accepts EA's precept that we should altruistically strive to positively influence the greatest number of people possible, and if most people who could exist will exist in the far future—"millions, billions, and trillions of years" from now[90]—then we should focus on how our present actions might affect these hypothetical future people rather than current people. Or, rather, we should focus on current people *only insofar* as doing so would help future people. As Benjamin Todd explains, "it might turn out that the best way to help those in the future is to improve the lives of people in the present, such as through providing health and education. The difference is that the major *reason* to help those in the present is to improve the long-term."[91] The claim isn't that future people matter more. It is, instead, a numbers game: there could be so many more future people that the expected value of

trying to help them may be orders of magnitude larger than the expected value of, e.g., lifting the 1.2 billion people in multi-dimensional poverty (as of 2026) out of that poverty.

Furthermore, it is important to recognize that longtermism, even in its moderate form, accepts the axiological theory of totalism.[92] This means that, if future people—the containers of value, or what Bostrom calls "value-structures"—on average will have lives of net-positive value, then *could exist* implies *should exist.* Longtermists aren't just interested in ensuring that future people have net-positive lives *conditional upon* those people existing; given their sympathies with totalist utilitarianism, they believe we have reason to bring as many future people into existence as cosmically possible.[93] This is precisely why longtermists (a) contend that there is "a *moral* case for space settlement,"[94] and (b) are so interested in calculating population sizes of *digital* people.[95] In theory, there could be far more digital people per 4-dimensional unit of spacetime than biological beings (the latter of whom would be living on the surface of terraformed exoplanets), and thus the goal must be to colonize the universe, build "planet-sized" computers powered by Dyson spheres, and run virtual-reality worlds full of trillions of sentient simulated minds. As Torres notes, colonizing space *itself* will almost certainly require our posthuman progeny to be digital, and hence the entire longtermist project is predicated on the possibility of digital sentience —i.e., on some form of functionalism or computationalism in the philosophy of mind being true.[96]

In many ways, longtermism could be seen as the apotheosis of the TESCREAL bundle. It combines the transhumanist goal of becoming or creating digital posthumans with the Cosmist eschatology of colonizing and reengineering the universe. It differs from Cosmism in providing an explicitly "ethical" foundation for this vision: the *reason* we must engage in "cosmic engineering" is because this is the best way to be effectively altruistic. Longtermists also agree with Rationalists that the Singularity could be one of the most important events in not just human but *cosmic* history, as ASI could enable us to realize a techno-utopian posthuman future marked by "astronomical" amounts of value—or erase it entirely if things go wrong.

One final point is worth making: the idea of "existential risk" is central to longtermism—indeed, to the entire TESCREAL movement. This was introduced in 2002 by Bostrom in an explicitly transhumanist context. He defined it as any event that would permanently prevent us from creating a transhumanist utopia. Examples of such risks include nuclear war, gray goo (self-replicating nanobots), runaway climate change, and an ASI takeover. Channelling the eugenic roots of TESCREALism, he also included "dysgenic pressures" on the list, arguing that fertility rates and "intellectual achievement" (as he understands it from a specifically white, male, Western perspective) are inversely related around the world. If dysgenic trends continue, whereby less "intellectually talented" individuals outbreed their more "talented" peers, then the overall "intelligence" of humanity could decline such that we are no longer able to develop the science and technology necessary to create utopia.[97] At the time that Bostrom was writing, the highest fertility rates were in African countries.[98]

The following year, Bostrom expanded his conception of "existential risk." Building on totalist utilitarianism, he characterized it as any event that would permanently prevent us from creating astronomical amounts of value in the future by colonizing space and building planet-sized computers to run simulated worlds populated by digital people.[99] This second definition, which is compatible with and complements the first, laid the

foundations for longtermism, and hence Bostrom could be considered the "Father" of this ideology. The idea was subsequently developed by Nick Beckstead, who would go on to become CEO of the FTX Future Fund, which channeled money from Bankman-Fried's crypto enterprise into longtermist research projects. In his 2013 PhD dissertation, Beckstead contended that the far future is of "overwhelming" moral importance, given how many people with "worthwhile" lives could—and therefore should—exist. To underline the point, he argued that we should prioritize saving the lives of people in rich countries over those in poor countries, all other things being equal. This is because people in rich countries are better positioned to influence the trajectory of civilizational development and, therefore, shape the very far future through contemporary actions.[100] Other longtermists, such as Toby Ord, have "enthusiastically" praised Beckstead's dissertation "as one of the most important contributions to the longtermist literature."[101]

Existential risks are thus fundamentally different from all other risks that do not threaten our "long-term potential" in the universe. As Bostrom writes, from a grand cosmic perspective, "a non-existential disaster causing the breakdown of global civilization is, from the perspective of humanity as a whole, a potentially recoverable setback: a giant massacre for man, a small misstep for mankind."[102] Elsewhere, he says this about the worst catastrophes in human history, including the AIDs pandemic, Black Plague, 1918 Spanish flu pandemic, and two world wars (including the Holocaust): "Tragic as such events are to the people immediately affected, in the big picture of things—from the perspective of humankind as a whole—even the worst of these catastrophes are mere ripples on the surface of the great sea of life" because "they haven't significantly affected the total amount of human suffering or happiness or determined the long-term fate of our species."[103]

In contrast, existential catastrophes would by definition determine our long-term fate and affect the total amount of happiness that could exist in our future light cone, and hence they are the "one kind of catastrophe that must be avoided at any cost."[104] This is why longtermists generally see existential risk mitigation as, to quote Bostrom once more, "priority number one, two, three and four," with the fifth priority being to colonize space as quickly as possible.[105]

## 3. Conceptualizing TESCREALism

Recall that the strong TESCREAL thesis states that we should conceptualize these ideologies as forming a single cohesive bundle across time, from the early 1990s to the present. Gebru and Torres substantiate this claim by pointing to historical, sociological, and philosophical considerations.

Historically, the "ESCREAL" ideologies all emerged directly out of the second-wave eugenics movement of modern transhumanism (the "T"). Extropianism was the first organized modern transhumanist movement. The two most notable singularitarians in the early 2000s were Ray Kurzweil and Eliezer Yudkowsky, both of whom participated in Extropianism through conferences or its mailing list. Modern Cosmism was introduced by the Extropian transhumanist and singularitarian sympathizer Ben Goertzel. Rationalism was founded by Yudkowsky to improve our odds of building an ASI that creates utopia rather than annihilating us, and EA was cofounded by Toby Ord, who years earlier had coauthored a paper with Bostrom defending transhumanism and then took a position at Bostrom's Future of Humanity Institute in 2006, founded in part to study and promote

radical human enhancement.[106] Finally, longtermism wove together many of the key ideas and themes of previous ideologies, having been introduced by the Extropian transhumanist, Nick Bostrom, and the younger EA, Nick Beckstead, who began posting on LessWrong in 2011 (and later became a Research Fellow at the Future of Humanity Institute, Program Officer for Open Philanthropy, and, as noted, CEO of the FTX Foundation funded by Sam Bankman-Fried).[107]

Sociologically, the communities that coalesced around each ideology in the acronym have overlapped considerably. Many EAs are also Rationalists and transhumanists.[108] Many longtermists write about the Singularity.[109] Many singularitarians envisage a future that is virtually identical to that advocated by Cosmists and longtermists.[110] And so on. Within all of these communities, the same group of people are widely lionized as luminaries with exceptional intelligence, rationality, and/or foresight, such as Nick Bostrom, Toby Ord, William MacAskill, Ray Kurzweil, Anders Sandberg, Robin Hanson, and Scott Alexander.[111] Within the accelerationist wing of the movement, figures such as Elon Musk, Gill Verdon, and Marc Andreessen have the same lofty status.

Furthermore, many of these groups are funded by the same wealthy individuals. Peter Thiel, for example, donated to the Machine Intelligence Research Institute (MIRI) and delivered the 2013 keynote address at the Effective Altruism Summit. Jaan Tallinn, a cofounder of Skype who has given talks at EA Global conferences, is a prominent longtermist who has financially supported TESCREAL groups like 80,000 Hours, AI Safety Camp, Rationalist Meetups, the Center for Applied Rationality, Lightcone Infrastructure, Rethink Priorities, Berkeley Existential Risk Initiative, Future of Life Institute, MIRI, Future of Humanity Institute, and Effective Altruism Funds (which, in turn, has supported organizations like the Global Priorities Project). The "main funders" of Open Philanthropy (OpenPhil) are Dustin Moskovitz, a Facebook cofounder, and his wife Cari Tuna.[112] OpenPhil has given money to Lightcone Infrastructure, Rethink Priorities, Berkeley Existential Risk Initiative, Future of Life Institute, MIRI, Future of Humanity Institute, Global Priorities Institute, and Effective Altruism Funds. Sam Bankman-Fried's FTX Future, run by Nick Beckstead, has similarly funded AI Safety Camp, Giving What We Can, Rethink Priorities, Berkeley Existential Risk Initiative, Lightcone Infrastructure, Longview Philanthropy, among others.[113] This is a brief glimpse of the tangled funding channels that connect a small number of millionaires and billionaires with a proliferation of TESCREAL-aligned organizations and institutes.

Philosophically, the TESCREAL movement is bound by similar epistemic commitments, such as the use of expected value theory, as well as moral inclinations, e.g., toward totalist utilitarianism. As Torres writes, "the emergence of these ideologies looks a lot like suburban sprawl, resulting in a cluster of municipalities without any clear borders between them—a conurbation of movements that share much the same ideological real estate."[114] Indeed, at the heart of this bundle is a techno-utopian eschatology in which we radically reengineer humanity, create a new posthuman species, colonize the universe, and spread the "light of consciousness" by ascending the Kardashev scale and establishing a sprawling multi-galactic civilization populated by many trillions of digital beings. But, as with many utopian ideologies, there is also an apocalyptic element: existential risks, discussed more below.

The TESCREAL movement is broadly libertarian, as noted earlier, even in the case of "doomers" who argue for the government to regulate AI, and one finds "an obsession with

'intelligence' and 'IQ' ... among TESCREAL advocates," with some boasting about their IQ, such as Yudkowsky, who has repeatedly described himself as a "genius" with an IQ of 143.[115] As Ruha Benjamin reminds us, "IQ is, above all, a eugenic concept, concocted to sort winners from losers and to justify the rules of game."[116] In 2023, a former EA reported that the Centre for Effective Altruism had tested a ranking system of community members called "PELTIV." The system added PELTIV points to those with IQs over 120 and subtracted them from those with IQs below 100.[117] Furthermore, Gebru and Torres note that some "leading figures in the TESCREAL community have approvingly cited, or expressed support for, the work of Charles Murray, known for his scientific racism." They also note that TESCREALists have repeatedly cited racist "notions of 'intelligence' that depend on IQ," a metric of "general intelligence" (the "GI" in "AGI") that was partly developed by 20th-century eugenicists to advance their racist, sexist, classist, ableist, and elitist visions of utopia. In his PhD dissertation (discussed below), Shane Legg—who introduced the term "AGI"—even "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.'"[118]

Zooming out, the TESCREAL movement as a whole is characterizable as promoting a cluster of interrelated "values" such as expansionism, colonization, optimization, maximization, quantification, and extractivism. These are core inclinations of the TESCREAL ideology, and its normative futurology can be seen as extending Western techno-capitalist and colonialist ideals into the stars.[119]

Taken together, these considerations are why Gebru and Torres defend the strong thesis according to which one should see these ideologies as a kind of package or bundle built around a particular strain of libertarian transhumanism that has become very influential within Silicon Valley.[120] It is to this latter issue—the emergence of AI companies out of the TESCREAL movement—that we now turn.

## 4. Origins of the AGI Race

Gebru and Torres advance two arguments: first, that the current race to build AGI is driven by the TESCREAL ideologies, and second, that this race and its TESCREAL-based "justification" is harmful and dangerous. This section will focus on the former claim, and the next section on the latter.

### *DeepMind*

The first major AI company with the explicit goal of creating AGI was DeepMind. This was cofounded in 2010 by Shane Legg, Demis Hassabis, and Mustafa Suleyman. Let's examine the extent to which these individuals have been involved in the TESCREAL movement.

Legg received his PhD in 2008 after completing a dissertation titled "Machine Super Intelligence." He then received $10,000 from the Canadian Singularity Institute for Artificial Intelligence. According to a 2023 interview, he read Ray Kurzweil's *Age of Spiritual Machines* in the early aughts and came to believe that a key part of Kurzweil's singularitarian view "was fundamentally right," namely, that "computation is likely to grow exponentially for at least three decades." He thus concluded that it would be possible to

train AI systems "on far more data than a human would experience in a lifetime. So, as a result of that, ... I predicted a 50% chance of AGI by 2028."[121]

In 2008, Legg had created a LessWrong account and responded to posts by Yudkowsky on AGI and the Singularity; the previous year, he exchanged emails on the SL4 mailing list, started by Yudkowsky in 2001, which describes itself as "a refuge for discussion of advanced topics in transhumanism and the Singularity."[122] According to *Wired*, "back in 2000, Yudkowsky came to speak at Goertzel's company," which Legg was working for. He "points to the talk as the moment when he started to take the idea of superintelligence seriously, going beyond the caricatures in the movies. Goertzel and Legg began referring to the concept as 'artificial general intelligence.'"[123] In 2010, Legg gave a talk at the Singularity Summit, founded by Yudkowsky, Kurzweil, and Thiel, on his AI research titled "Measuring Machine Intelligence."[124]

At the same Singularity Summit, Hassabis—who met Legg in the Gatsby Computational Neuroscience Unit of University College London—delivered a talk exploring "a systems neuroscience approach to building AGI."[125] After the event had ended, Hassabis followed Thiel back to his mansion and solicited funding to start DeepMind. Thiel, himself a TESCREAList, agreed and gave Hassabis $1.85 million to start the company.[126] Later, when Google acquired the company in 2014, "Thiel's venture capital firm, Founders Fund, owned more shares than all three of DeepMind's co-founders."[127] Other DeepMind investors included Elon Musk and Jaan Tallinn, both of whom have funded TESCREAL organizations; Tallinn also formerly served as DeepMind's director.[128]

In 2011, Hassabis gave a talk about AI at the Future of Humanity Institute, and DeepMind subsequently included Bostrom as a member of the company's "Ethics and Society" team.[129] When the Musk-funded Future of Life Institute hosted a conference on AI, Hassabis joined Tallinn, Musk, Kurzweil, and Bostrom on stage to discuss the promises and dangers of machine superintelligence. In fact, the "Time 100" entry for Hassabis in *Time* magazine was written in 2017 by Kurzweil.

Meanwhile, Suleyman's allegiances are less well known. During an 80,000 Hours podcast interview, he averred that he has "long been a fan of the podcast and the [EA] movement," and later stated that he, along with others at DeepMind and its competitor OpenAI, are members of the "AI Safety" community.[130] The field of AI Safety directly emerged out of the TESCREAL movement, as explained in subsection 5.1.

From its inception, then, DeepMind has been closely linked to the TESCREAL movement.

### OpenAI

OpenAI was cofounded in 2015 by Sam Altman, Elon Musk, Ilya Sutskever, Greg Brockman, and others. It began with $1 billion in funding from Musk (after he had invested in DeepMind), Peter Thiel, and additional investors.[131] Jaan Tallinn "offered to financially support OpenAI's safety research and met regularly with [Dario] Amodei and others at the organization."[132] In 2023, Andreessen Horowitz was among several venture capital firms that collectively invested over $300 million into OpenAI. Andreessen Horowitz was cofounded by Marc Andreessen, a leading advocate of effective accelerationism who, in 2023, included "TESCREAList" in his bio on Twitter (subsequently rebranded as X).[133] Another $30 million was later invested by OpenPhil, largely funded by Dustin Moskovitz.[134]

Altman has extensive connections to the TESCREAL movement. According to a *New York Times* profile, he is "the product of a strange, sprawling online community that began to worry, around the same time Mr. Altman came to the Valley, that artificial intelligence would one day destroy the world. Called rationalists or effective altruists, members of this movement were instrumental in the creation of OpenAI."[135] Another profile describes Altman as having "embraced the techy-catnip utilitarian philosophy of effective altruism."[136] Altman credits Eliezer Yudkowsky as having inspired him and others in the field to pursue AGI, writing that Yudkowsky "got many of us interested in AGI, helped DeepMind get funded at a time when AGI was extremely outside the Overton Window, [and] was critical in the decision to start OpenAI, etc."[137] In 2015, he wrote on his personal blog that "Bostrom's excellent book *Superintelligence* is the best thing I've seen on this topic. It is well worth a read."[138] In early 2025, he posted on X that he "always wanted to write a six-word story. Here it is: Near the Singularity; unclear which side."[139]

Altman is also a transhumanist who invested $180 million into Retro Biosciences, a longevity research company. And he believes that our brains will be digitized within our lifetimes. In 2018, he was one of 25 people to sign up with Nectome, a startup offering to preserve people's brains so they can be uploaded to computers.[140] During a social media exchange with Yudkowsky, he agreed that "future galaxies are indeed at risk" when it comes to getting AGI right, and says, "I do not believe we can colonize space without AGI."[141,142] He also echoes TESCREALists like Bostrom in suggesting that the outcome of advanced AI, which he dubs the "magic intelligence in the sky," will likely either be "lights out for all of us" or a paradisiacal world so marvelous that one begins to "sound like a really crazy person" when talking about it.[143] In another interview, he says that AGI will "most likely sort of lead to the end of the world, but in the meantime there will be great companies created with serious machine learning."[144]

These statements are worth pausing on. The idea that the outcome of superintelligence will likely be binary—either annihilation or utopia—is ubiquitous within the TESCREAL movement. It dates back at least to I. J. Good's work in the late 1950s and 1960s; he argued that since "an ultraintelligent machine could design even better machines[,] there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind."[145] He added that "whether this will lead to a Utopia or to the extermination of the human race will depend on how the problem is handled by the machines."[146] Musk channels this thinking when he says that "the most likely outcome" of superintelligence "is awesome. … But I think it's either going to be super-awesome or super-bad. It's probably not going to be something in the middle." He gives annihilation a 20% probability, and utopia an 80% probability.[147]

Indeed, this points to a crucial link between the TESCREAL movement and the AGI race. On the one hand, if we successfully design a "value-aligned" ASI—that is, an AGI that is controllable by humanity or some subset of humans—then we can delegate it the task of "paradise-engineering" (Bostrom 2020). We get to become "semi-mortal uploaded creatures with Jupiter-sized minds," in Bostrom's words.[148] But, on the other hand, if this fails, ASI will almost certainly destroy humanity and, along with us, the utopian future of endless delights that we could have otherwise created. As Bostrom writes, the "default outcome" of misaligned superintelligence is "doom."[149]

Both possibilities account for why the AGI race emerged out of the TESCREAL movement: the seductive reward of the first possibility is obvious, and many TESCREALists

believe that building utopia will be impossible without ASI. Hence, we have a kind of moral obligation to build ASI as soon as possible. To quote Bostrom once again, "all the plausible paths to a really great future involve the development of machine superintelligence at some point."[150] With respect to the second possibility, many leading figures in the AGI race, including Altman, believe that *their* company is best positioned to ensure a utopian outcome. This is why there has been a proliferation of companies over the past 15 years—DeepMind, OpenAI, Anthropic, xAI, and so on. Each sees itself as more responsible than the others, and each is thus racing to reach the AGI finish line before everyone else. Musk, for example, was compelled to cofound OpenAI with Altman because he saw Hassabis as "a supervillain who needed to be stopped," and "would make unequivocally clear that OpenAI was the good to DeepMind's evil."[151] Anthropic, examined below, was founded because people lost faith that Altman was the right person to bring about AGI.

Even more, there is also the specter of rogue states building AGI, which could enable them to establish a global totalitarian regime. As Altman wrote to his employees, "if an authoritarian government builds AGI before we do and misuses it, we will have also failed at our mission" at OpenAI, and thus "we almost certainly have to make *rapid technical progress* in order to succeed at our mission."[152]

These considerations are why OpenAI and the other companies were founded: to bring about utopia as quickly as possible while simultaneously containing the existential risks of the technology, which could be greatly exacerbated if other companies or states were to develop it first.

### *Anthropic*

Anthropic was started in 2021 by seven former employees at OpenAI, including the siblings Daniela and Dario Amodei. The latter is a noted EA and longtermist who has had an account on LessWrong since at least 2008.[153] Like Shane Legg, he "first [became] interested in AI after reading Ray Kurzweil's *The Singularity Is Near: When Humans Transcend Biology*, which predicted that AI would reach human intelligence by 2029 and that people would merge with machines by 2045."[154] While all of the leading AI companies in the West have been heavily influenced by the TESCREAL ideologies, Anthropic stands out for its deep alignment with EA-longtermist principles. As a *New York Times* article puts it,

> all of the major AI labs and safety research organizations contain some trace of effective altruism's influence, and many count believers among their staff members. ... No major AI lab embodies the EA ethos as fully as Anthropic. Many of the company's early hires were effective altruists, and much of its start-up funding came from wealthy EA-affiliated tech executives.[155]

As with OpenAI, the impetus behind its founding concerned both possibilities mentioned in the previous subsection—i.e., utopia and an AI catastrophe. Dario Amodei "and the other Anthropic founders," notes Karen Hao, built "up their own mythology about Anthropic, not OpenAI, [being] a better steward of what they saw as the most consequential technology."[156] Amodei and other researchers were dissatisfied with OpenAI's commitment to AI safety, and hence established a competitor in hopes of reaching the AGI finish line first —that is, with their *safer* version of AGI than what OpenAI or DeepMind would create.

Consistent with this, wealthy TESCREALists worried about AI risk were major funders of the company. An initial investment of $124 million was led by Jaan Tallinn and Dustin Moskovitz, among others, with $25 million being provided by Tallinn.[157] One year later, Sam Bankman-Fried "led the $580 million Series B venture capital round for Anthropic."[158] Once again, the founders and funders have both been firmly rooted in the TESCREAL movement.

### *xAI*

xAI was founded in 2023 by Elon Musk. It boasts of having the largest AI supercomputer in the world, as of this writing, named Colossus.[159] Musk, like Shane Legg, Sam Altman, and Dario Amodei, has extensive links to the TESCREAL movement. He is a transhumanist whose company Neuralink is trying to merge the human brain with AI, enabling us "to save and replay memories. … Ultimately, you could download them into a new body or robot body."[160] In early 2025, he posted on his social media website that "we are on the event horizon of the singularity."[161] He promoted William MacAskill's 2022 book promoting longtermism, titled *What We Owe the Future*, writing that longtermism "is a close match for my philosophy."[162] In 2022, he exchanged private messages with MacAskill about purchasing Twitter, with Sam Bankman-Fried. MacAskill introduced the two by saying that "you both have interests in games, making the very long-run future go well, and buying Twitter. So I think you'd have a good conversation!" MacAskill then boasted that Bankman-Fried is "moving $100M-$1B this year to improve the future of humanity." When Musk asked "You vouch for him?," MacAskill responded with "Very much so!"[163] Musk has consistently characterized his motive for purchasing Twitter in longtermist language; the reason he bought the platform is "because it's important to the future of civilization," elsewhere declaring that "this is a battle for the future of civilization. If free speech is lost even in America, tyranny is all that lies ahead."[164] Since "tyranny" would impede the realization of techno-utopia, controlling Twitter, the most politically important social media platform, is paramount.

In 2022, Musk retweeted an article by Nick Bostrom titled "Astronomical Waste." The original tweet included the line: "Likely the most important paper ever written."[165] This article outlines a moral case for why (a) mitigating existential risk should be our top four global priorities, and (b) we should colonize the universe as soon as possible, build "planet-sized" computers around other stars, and run virtual-reality worlds full of, according to Bostrom, some 10^38 digital people per century in the Virgo Supercluster alone (this was his earlier estimate).[166] Echoing the totalist utilitarian pillars of longtermism, Musk says that "what matters … is maximizing cumulative civilizational net happiness over time," and that "we have a duty to maintain the light of consciousness, to make sure it continues into the future."[167]

As with the other companies above, the explicit goal of xAI is to build AGI, or what Musk describes as "basically a digital God."[168] Musk has argued that AGI could be "more dangerous than nukes," but, as noted earlier, he also believes that the most likely outcome of AGI will be utopia.[169] According to Torres, most of Musk's companies only make sense when viewed through the lens of TESCREALism.[170] For example, Neuralink hopes to "kickstart transhuman evolution with 'brain hacking' tech."[171] Tesla "makes cars but is really an AI company working on computer vision, image recognition, machine learning and

autonomous decision making."[172] SpaceX aims to fulfill a key aspect of the Cosmist and longtermist projects, namely, spreading beyond Earth to colonize the universe; its headquarters even include a conference room named after Konstantin Tsiolkovsky, a pioneer of astronautics who helped found Russian Cosmism.[173] And xAI is trying to build "safe" superintelligence before other companies reach the finish line, thus catalyzing the Singularity.

In service of this end, xAI has hired individuals from the TESCREAL community to advise its research, such as Dan Hendrycks. In addition to working for xAI, Hendrycks is "the executive and research director of the Center for AI Safety, which was awarded a grant of US$5,160,000 from Open Philanthropy."[174] In a post on the Effective Altruism Forum, he reports that he "was advised … to get into AI to reduce x-risk [i.e., existential risk], and so settled on this rather than proprietary trading for earning to give," as Bankman-Friend initially did when he took a job at Jane Street Capital.[175]

## 5. Critiquing TESCREALism

Torres and Gebru argue that the race to build AGI is both harmful and dangerous. This section examines several arguments in support of this claim, outlined in their coauthored paper as well as in separate publications and presentations.

### *Is Safe "AGI" Possible?*

The TESCREAL movement can be roughly divided into two main camps: the *accelerationists* (or "effective accelerationists," abbreviated as "e/acc") who believe that the probability of an existential catastrophe if AGI is built in the near future is very low, and the *doomers* who believe that the probability is very high. Both groups accept the same eschatological vision, and both believe that we should build AGI as quickly as possible.[176]

Because the risks are low, accelerationists argue that we should radically accelerate AI development and eliminate government regulation. They would say that, insofar as there *is* any risk associated with AGI, the best countermeasure is the free market: let a million AGIs bloom, since the best way to stop a bad AGI is with a good AGI.[177] Hence, the more AGIs there are (by open-sourcing AI models, cutting regulation, etc.), the greater the chance that the good AGIs will neutralize the bad ones. This "move fast" position is closely related to the Proactionary Principle defended by Max More, and indeed the e/acc TESCREAList Marc Andreessen has argued that those who impede the development of AI are no better than murderers, given how many lives AI will supposedly save.[178]

In contrast, AI doomers tend to favor top-down government regulation, thus rejecting the free-market solution to mitigating AI risk. Their view is slightly more nuanced than the accelerationists'. One can understand it by distinguishing between *AI capabilities research* and *AI safety research*. The former aims to build AGI, whereas the latter aims to ensure that the AGI we build is controllable or "value-aligned." Doomers argue that if capabilities research leads the way over safety research, then the probability of total annihilation is very high. As Yudkowsky writes, "the most likely result of building a superhumanly smart AI, *under anything remotely like the current circumstances*, is that literally everyone on Earth will die."[179]

However, if safety research leads the way, then the most likely outcome will be utopia. From this perspective, we can see how Yudkowsky and other doomers are *not* anti-AGI—to the contrary, they want AGI no less than the accelerationists, but only if safety research leads the way over capabilities research. As the "2024 Communications Strategy" report published by Yudkowsky's MIRI puts it, "we remain committed to the idea that failing to build smarter-than-human systems someday would be tragic and would squander a great deal of potential. We want humanity to build those systems, but only once we know how to do so safely."[180] This point was nicely encapsulated by a social media exchange involving Rob Bensinger, a long-time MIRI employee. Someone asked whether "there is an 'e/acc' for everything but … AI-enabled warfare and authoritarian surveillance." "Last I checked," Bensinger replied, "the term for that kind of e/acc is 'doomer.'"[181]

Contrasting with both the accelerationists and doomers, Gebru and Torres argue that the very *attempt* to build AGI is "*inherently* unsafe" because what they describe as "AGI" would be an "unscoped" system—an "everything machine" that is imagined to excel on complex tasks in every cognitive domain of interest.[182] A "scoped" system has well-defined functions, which can be tested in relevant operational conditions. For example, evaluating the "audio circuitry for devices such as laptops" involves

> drop testing, constantly dropping devices to understand the manner in which their functionality degrades when they are exposed to shocks, placing the devices in extremely cold or hot environments …, frequently restarting them, and performing different types of tests to understand the behavior of these systems under conditions that they were not normally meant to operate in.

Gebru and Torres add that "these stress tests occurred in addition to extensive testing and documentation under conditions that the devices were meant to be operational."[183] Heidy Khlaaf makes a similar point, writing that "the lack of a defined operational envelope for the deployment for general multi-modal [AI] models has rendered the evaluation of their risk and safety intractable, due to the sheer number of applications and, therefore, risks posed."[184]

Since there is no way to effectively test an "everything machine" that excels at all possible tasks of interest, there is *no way to be certain* that the type of system AGI proponents claim to be building would be safe.[185] Worse, there isn't even any clear conception of what AGI *is* in the first place, as different theorists and companies define the term in nontrivially different ways. Even OpenAI's website contains inconsistent definitions, as when it describes AGI as "AI systems that are generally smarter than humans" in one article,[186] and "highly autonomous systems that outperform humans at most economically valuable work" in another.[187] Others define it as systems that "can match or exceed the cognitive abilities of human beings across any task" (IBM),[188] "AI that's at least as capable as humans at most cognitive tasks" (DeepMind),[189] and "a universal algorithm for learning and acting in any environment" (Russell and Norvig).[190] Dario Amodei takes a different approach, preferring the term "powerful AI" to "AGI," which he defines as having properties like being "smarter than a Nobel Prize winner across most relevant fields," "not hav[ing] a physical embodiment," and being able to "absorb information and generate actions at roughly 10x-100x human speed."[191]

Given that none of these companies or theorists can agree about what AGI is, and that trying to build such an unscoped system is inherently dangerous, Gebru and Torres argue that we should eschew projects to build such systems—the goal of AI capabilities research. Instead, we should opt to develop "narrow AI" tools that "might specifically be trained to identify certain types of plant disease … or perform machine translation in specific language," and so on, since these systems "have task definitions and expected inputs and outputs for which appropriate tests can be created and results can be compared to expected behavior."[192] The dream of "safe" AGI—the goal of AI safety research—is a fantasy. There isn't even *such a thing as* "AGI" to build toward, as the concept of AGI is inherently unscoped, ill-defined, vague, and unruly.

This conclusion, however, is anathema to TESCREALists because they see AGI as necessary for creating utopia. As noted earlier, all paths to utopia involve building superintelligence.[193] Elsewhere, Bostrom writes that "because of the way we have defined existential risks, a failure to develop technological civilization would imply that we had fallen victims of an existential disaster," where AGI/ASI is integral to developing a technologically advanced posthuman civilization.[194] Hence, failing to develop AGI/ASI would constitute an existential catastrophe. This is precisely why the field of AI safety emerged out of the TESCREAL movement: there is only one way forward, according to TESCREALists, so we should proceed with caution by examining the ways that AGI might backfire in order to neutralize the risks. The goal is, as it were, to keep our technological cake and eat it, too.

Gebru and Torres suggest an alternative strategy: since there is no way to ensure that AGI — if such a thing could even exist — will be safe, we should refuse to build it in the first place. For this reason, their position is fundamentally at odds with both the accelerationist and doomer camps within the TESCREAL movement.

### *The Dangers of TESCREALism*

Not only is safe AGI a mirage, but the techno-utopian beliefs of TESCREALists enable them to ignore or downplay the harms caused by racing to build AGI. Consider the following line of reasoning:

> (1) The large language models (LLMs) that power frontier models like OpenAI's ChatGPT, DeepMind's Gemini, Anthropic's Claude, and xAI's Grok are the stepping stones to AGI.
> (2) Once we have AGI, ASI will soon follow (e.g., due to recursive self-improvement).
> (3) If ASI is safe (i.e., is controllable or value-aligned), then it will usher in a utopian paradise of immortality, perfect happiness, radical abundance, space colonization, and "astronomical" amounts of "value."[195] In Bostrom's words, it will produce a "solved world" in which every problem has been solved forever.[196]

In numerous publications since 2021, Torres has argued that the marriage of utopianism and utilitarianism is extremely dangerous. If the ends can morally justify the means, and if the ends are a literal utopia marked by astronomical value, then what exactly is off the table

for realizing this end?[197] Hence, if ASI is the key to utopia, and if current LLMs are the stepping stones to ASI (via AGI), then every harm caused by these LLMs can be "justified" by the utopian aim of reaching ASI. Let's examine this in more detail, dividing the dangers of this way of thinking into two categories.

*Trivializing the Harms of AI*

The first category of danger could be classified as "passive." It concerns the way that TESCREALism inclines adherents to ignore, dismiss, and trivialize the actual harms caused by LLMs. These harms include worker exploitation (especially in the Global South), intellectual property (IP) theft, using valuable resources like fresh water, the growing carbon footprint of generative AI, perpetuating noxious stereotypes through biased algorithms, mental health problems caused by AI chatbots, as well as the production and propagation of deepfakes, disinformation, misinformation, and propaganda. OpenAI, for example, hired a company that paid workers in Kenya as low as $1.32 an hour to label horrific material to train their LLMs, resulting in some workers being diagnosed with PTSD.[198] Are such harms justifiable?

They can be, from the perspective of TESCREALism, because these LLMs will get us to AGI, which will get us to ASI and then utopia, if designed properly. Recall from earlier that Bostrom describes the worst atrocities of the 20th century as "mere ripples on the great sea of life," since they haven't affected the total amount of "value" that could exist in the universe. Elsewhere, he says that giant massacres for man may be small missteps for mankind, so long as these massacres are non-existential in nature. Once again, existential catastrophes matter *far more* than non-existential catastrophes because they would, by definition, prevent us from creating a posthuman utopia among the stars, whereas non-existential catastrophes wouldn't. This is a difference of kind rather than degree.[199] Since all of the harms mentioned above are non-existential, they are nothing but mere ripples or minor missteps on the road to paradise; mitigating them should therefore not be a global priority for humanity.[200] Bostrom drives the point home in writing:

> Unrestricted altruism is not so common that we can afford to fritter it away on a plethora of feel-good projects of suboptimal efficacy. If benefiting humanity by increasing existential safety achieves expected good on a scale many orders of magnitude greater than that of alternative contributions, we would do well to focus on this most efficient philanthropy.[201]

Caring about how AI companies harm people in the Global South, steal from writers and artists, exacerbate the climate crisis, etc. would count on Bostrom's view as "feel-good projects of suboptimal efficacy," because the associated harms aren't existential.

Geoffrey Hinton holds the same view as Bostrom. He dismisses concerns about the concrete harms of LLMs, raised by figures like Gebru, as not being "as existentially serious as the idea of these [AI models] getting more intelligent than us and taking over."[202] Yudkowsky similarly describes algorithmic bias as a "short-term and small" issue when compared to what awaits: the "glorious transhumanist future."[203] He says that if AI ethicists, a group that includes Gebru and Torres,

would leave the people trying to prevent the utter extinction of all humanity alone I should have no more objection to them than to the people making sure the bridges stay up. If the people making the bridges stay up were like, "How dare anyone talk about this wacky notion of AI extinguishing humanity. It is taking resources away that could be used to make the bridges stay up," I'd be like "What the hell are you people on?" Better all the bridges should fall down than that humanity should go utterly extinct.[204]

The TESCREAL worldview thus inclines adherents to minimize or trivialize all non-existential harms. In this way, it provides a superficially plausible "moral" excuse for pursuing AI capabilities and safety research without regard to the real-world consequences for marginalized peoples, writers and artists, the environment, etc. This is not to say that TESCREALists would *deny* the significance of these harms in *absolute* terms. The point is that *relative* to the disvalue of delaying or never reaching utopia, these harms, as Torres puts it, are but "molecules in a drop in the ocean."[205] They simply don't matter much in the grand scheme of things, and hence if some people get trampled by the march of progress, we shouldn't spend more than a moment lamenting it. As Hilary Greaves and William MacAskill write, "every \$100 spent" on AI safety research would, by increasing the probability of utopia, have "an impact as valuable as saving one trillion … lives."[206] Utopianism combined with a utilitarian moral calculus thus suggests that one does more good in the world by donating a mere \$100 to AI safety than by saving 999 billion human lives. This leads us to the second category.

*Justifying Extreme Actions*

The marriage of utopianism and utilitarian ethics could also justify extreme measures, actions, or interventions to "protect" and "preserve" our "vast and glorious" future among the stars.[207] Though not discussed in Gebru and Torres (2024), Torres articulated this point in a 2021 critique of longtermism, writing that "elevating the fulfilment of humanity's supposed potential above all else could nontrivially increase the probability that actual people—those alive today and in the near future—suffer extreme harms, even death."[208] Consider Bostrom's claim (echoing Greaves and MacAskill above) that if there is even just "a mere 1 per cent chance" of $10^{54}$ people existing in our future light cone, then "the expected value of reducing existential risk by a mere *one billionth of one billionth of one percentage point* is worth a hundred billion times as much as a billion human lives" today.[209] This calculation could yield a justificatory argument, based on expectational utilitarianism, for engaging in extreme actions, even violence, mass murder, and genocide, for the sake of the far future. "It is simply too reminiscent," writes Olle Häggström in reference to Bostrom's claim, "of the old saying 'If you want to make an omelet, you must be willing to break a few eggs,' which has typically been used to explain that a bit of genocide or so might be a good thing, if it can contribute to the goal of creating a future utopia."[210] Häggström goes on to outline a scenario in which a US president following Bostrom's reasoning might launch a barrage of nuclear weapons at Germany to prevent a low-probability existential catastrophe, thus annihilating the country.[211]

A similar example comes from Peter Singer, a noted EA and eugenicist who once defended longtermism in an article coauthored with Nick Beckstead.[212] Referencing

Torres's work, Singer pushes back against the TESCREAL notion of utopia. He warns that "viewing current problems through the lens of existential risk to our species can shrink those problems to almost nothing, while justifying almost anything that increases our odds of surviving long enough to spread beyond Earth." He continues: "I am not suggesting that any present exponents of the hinge of history idea"—the view that we live during a period of excessively high existential risk, largely due to the prospect of unaligned AGI—"would countenance atrocities. But then, Marx, too, never contemplated that a regime governing in his name would terrorize its people."[213] The danger, in other words, lurks within the ideologies themselves; so long as such ideologies have their ideologues, there remains the risk of people acting on them.

Singer is correct that there were no TESCREALists calling for violence, conflict, or the like when he published his article in 2021, but this is no longer the case. Many TESCREAL doomers believe that we are now in an apocalyptic moment due to the rapid advancement of AI capabilities and the comparatively slow progress of AI safety. In late 2022, an AI safety workshop held in Berkeley, California, where Yudkowsky's MIRI is located, explored strategies for slowing down or temporarily halting capabilities research so that safety research can catch up. Notes from the workshop included proposals like: "Solution: be Ted Kaczynski" and "start building bombs from your cabin in Montana and mail them to DeepMind and OpenAI, lol." Another line read: "Strategy: We kill all AI researchers."[214] This event was funded by FTX Future Fund, run at the time by Nick Beckstead, and was organized by three people: one was "a guest manager at Effective Altruism Funds" whose work had been funded by OpenPhil;[215] another worked for Dan Hendryck's Center for AI Safety and the Center on Long-Term Risk, and later became an AI Fellow in the US House of Representatives;[216] the third organizer subsequently worked for MIRI.[217] These workshop notes, quoted above, were leaked to Torres after Torres publicly shared threats of physical violence from the TESCREAL community, which were sent in response to their critiques.[218]

The following year, in 2023, Yudkowsky published an op-ed in *Time* magazine that argued for states engaging, if necessary, in military strikes targeting "rogue datacenters," even at the risk of triggering a devastating nuclear war. The reasoning is that an all-out nuclear exchange probably wouldn't result in an existential catastrophe, thus permanently foreclosing the realization of utopia. Some studies corroborate this conclusion, finding that a full-scale conflict that exhausts the nuclear stockpiles of Russia and the US could kill "more than 5 billion" people.[219] If 5 billion were to perish, it would leave a reassuring ~3 billion to rebuild global modern civilization. In contrast, Yudkowsky believes that if capabilities research continues, it will result in a misaligned AGI that destroys humanity and, along with us, the "glorious transhuman future" that awaits. Hence, he concludes that nuclear war is worth risking to prevent AGI from being built in the near future—that is, until safety research catches up.[220] When he was asked on social media, "How many people are allowed to die to prevent AGI?," he responded that "there should be enough survivors on Earth in close contact to form a viable reproductive population, with room to spare, and they should have a sustainable food supply. So long as that's true, there's still a chance of reaching the stars someday."[221] Estimates of the viable human population range widely from 150 to around 40,000 people, meaning that well over 8 billion people could perish without irreversibly foreclosing our "chance of reaching the stars someday."[222]

The dangers examined in this subsection could be classified as "active" since they involve people finding themselves amid what they perceive to be a do-or-die apocalyptic moment, causing harm through violence or extreme measures by acting on their utopian beliefs, and believing they are "justified" in such actions by the means-ends reasoning of utilitarianism. In both the passive and active cases, the danger arises from the amalgam of utopianism and utilitarianism, which can enable virtually any *action or non-action* to become morally acceptable—indeed, *obligatory*—given the astronomical stakes of utopia. As Torres writes, "over and over again throughout history, the combination of these two ingredients—utopianism and the belief that ends justify the means—has been disastrous."[223] If this combination has been disastrous in the past, we should worry that it will be disastrous in the future, too.

### Who Is Utopia For?

The issues discussed in "Trivializing the Harms of AI" and "Justifying Extreme Actions" concern dangers associated with the *march to utopia*. But what about utopia *itself*? One of the most striking features of the TESCREAL literature is the absence of any serious consideration of what the future should look like from perspectives falling outside the TESCREAL tradition. Recall that this tradition is marked by a commitment to ideas or values like eugenics, libertarianism, capitalism, expansionism, colonization, extractivism, quantification, maximization, optimization, and utilitarianism. The techno-utopian visions at the core of TESCREALism were crafted almost entirely by white men in Silicon Valley and at elite institutions like the University of Oxford. Hence, these visions reflect the particular social privileges, ideological commitments, and normative preferences of their authors. The TESCREAL movement does not, it seems, aim to bring about an inclusive future for everyone through democratic means; if it did, one would expect TESCREALists to include representatives of different perspectives at the table of futurological debate, such as the perspectives of feminism, Queerness, Disability, Islam, Buddhism, Indigenous cultures, and various other non-Western thought traditions, to name a few.[224] The aim instead is to impose the TESCREAL vision on everyone else, with or without their consent. If safe AGI is even possible, it would provide a powerful means of doing this, since it would presumably be controlled by the AI company or companies that build it, and we saw in earlier sections that major AI companies like DeepMind, OpenAI, Anthropic, and xAI directly emerged out of the TESCREAL movement. In other words, if safe AGI were built, it would be controlled by people aligned with the TESCREAL worldview.

This raises the question: if the TESCREAL movement were to succeed in bringing about utopia through AGI—which Gebru and Torres suggest is impossible—who would this utopia be for? The concept of utopia is inherently exclusionary: someone is always left out; if no one were left out, then it wouldn't be utopia. As Monika Bielskyte puts it, utopia involves "a kind of eugenic elimination" of certain groups deemed to be undesirable.[225]

Torres argues that the utopia of TESCREALism would exclude much of the vast diversity of human cultures, traditions, religions, and peoples. It would exclude those with certain disabilities (e.g., intellectual disabilities), and would erase much of what makes our species so exquisitely unique and special within the known universe. Robert Sparrow further notes that if transhumanism—and hence TESCREALism—were put in practice, the outcome would be largely indistinguishable from what the first-wave eugenicists hoped to

bring about. In his words, "the ultimate conclusions of the new eugenics are remarkably similar to those of the old."[226]

It is, therefore, not clear that most of humanity would have a home in the "utopian" future envisaged by TESCREALism. If such people *were* welcome in the future, one would expect the TESCREAL movement to have taken seriously the rich array of futurological visions articulated by those outside the TESCREAL community, which it has not.

Indeed, it is not even clear that the natural world, teaming with nonhuman organisms and bustling ecosystems full of wonders still unknown to modern science, would have any place in utopia. MacAskill, for example, argues that our systematic obliteration of the biosphere may be net positive. This is because "if we assess the lives of wild animals as being worse than nothing, which I think is plausible … then we arrive at the dizzying conclusion that from the perspective of the wild animals themselves, the enormous growth and expansion of *Homo sapiens* has been a good thing."[227] To put this in perspective, the 2022 Living Planet Report finds that the global population of all wild vertebrates has declined by a staggering 69% since 1970, due largely to human expansion around the globe and the attendant pollution, habitat fragmentation, and global warming that it has caused.[228] MacAskill thus suggests that we should celebrate this precipitous decline in global biodiversity. After all, once the biological world has been entirely decimated and superseded by a new digital world, we could simulate natural ecosystems while ensuring that the digital flowers, insects, and woodland creatures have lives that are not "worse than nothing."

Torres thus concludes that TESCREALists are presenting a deeply problematic set of options that are exclusivist, undemocratic, and dystopian. One way to make the point goes like this, which accepts the TESCREALists' framing: if AGI is uncontrollable, then *all* of humanity loses, because everyone will likely die. However, if AGI is controllable, then *most* of humanity loses, because there is no reason to believe that the world's many cultures, traditions, peoples, etc. would have a home in the utopia of TESCREALism. Most people around the world do not even have a say in what we should be aiming for in the future and with AI.[229,230] This is a lose-lose situation for most of humanity, and hence the entire TESCREAL project should be rejected.

### *Pro-Extinctionism*

It might not just be marginalized communities and the natural world that have no place in utopia—our species itself would likely be excluded. This leads Torres to argue that TESCREALism is best seen as a *pro-extinctionist* ideology, if only in practice. Rather than being a fringe view embraced by some philosophical pessimists, radical environmentalists, and negative utilitarians, pro-extinctionism has in fact become very influential within powerful sectors of Silicon Valley.

To understand this, Torres notes that there are different types of human extinction that our species could undergo. The most relevant are what they call "terminal" and "final" extinction. *Terminal extinction* would occur if and only if our species were to disappear entirely and forever. *Final extinction* would occur if and only if our species were to disappear entirely and forever *without* leaving behind any successors. The latter subsumes the former while adding an additional condition. Final extinction thus entails terminal extinction, but not vice versa.

This yields two kinds of pro-extinctionism: one specifically aims to bring about final extinction. This is what the aforementioned philosophical pessimists (e.g., Eduard von Hartmann), radical environmentalists (e.g., the Voluntary Human Extinction Movement), and negative utilitarians (e.g., Efilists) strive to bring about through means like antinatalism (refusing to have children) or omnicide (the murder of all people).[231] The other specifically aims for terminal extinction *without* final extinction. For example, if our species were to create posthuman successors that replace us such that *Homo sapiens* ceases to be, we will have undergone terminal but not final extinction. This is precisely where this kind of pro-extinctionism overlaps with digital eugenics—the eugenics of transhumanism—since the goal of terminal extinction would coincide with the emergence of a "superior" new species of posthuman entities.[232] In other words, the form that eugenics takes in the TESCREAL bundle *is pro-extinctionism*.

Here are some examples of pro-extinctionist thought within the broad TESCREAL movement. We begin with Hans Moravec, a keynote speaker at the Extropy Institute's first EXTRO conference who has influenced the development of TESCREALism. In 1988, he described himself as "an author who cheerfully concludes that the human race is in its last century, and goes on to suggest how to help the process along."[233] Larry Page, cofounder of Google, which acquired DeepMind in 2014, asserts that "digital life is the natural and desirable next step in ... cosmic evolution and that if we let digital minds be free rather than try to stop or enslave them, the outcome is almost certain to be good."[234] Derek Shiller, a researcher at Rethink Priorities, funded by Jaan Tallinn and OpenPhil, writes that

> it is plausible that in the not-too-distant future, we will be able to create artificially intelligent creatures with whatever physical and psychological traits we choose. Granted this assumption, it is argued that we should engineer our extinction so that our planet's resources can be devoted to making artificial creatures with better lives.[235]

Yudkowsky reports that he is worried not "about being replaced by a better organism," only that the organism replacing us might not "be better."[236] In other words, he would not object to a posthuman species replacing humanity so long as it is superior (which is precisely the constraint imposed by digital eugenics). Elsewhere, he declares:

> If sacrificing all of humanity were the only way, and a reliable way, to get ... god-like things out there—superintelligences who still care about each other, who are still aware of the world and having fun—I would ultimately make that trade-off.[237]

We are *not* currently faced with this trade-off, he adds. But if we were, he would sacrifice our species for the sake of AI. These remarks are from an interview conducted by the Founder of Emerj Artificial Intelligence Research, Daniel Faggella, who holds a similar view. He argues that "the great (and ultimately, only) moral aim of artificial general intelligence should be the creation of [a] Worthy Successor—an entity with more capability, intelligence, ability to survive and ... moral value than all of humanity." In another document, he defines "Worthy Successor" as "a posthuman intelligence so capable and morally valuable that you would gladly prefer that it (not humanity) control the

government, and determine the future path of life itself."[238] Of note is that Faggella hosted an event in 2025 at a San Fransisco mansion. The event was titled "Worthy Successor: AI and the Future after Humankind."[239] It reportedly included "a star-studded guest list," including "team members from OpenAI, Anthropic, DeepMind, and other AGI labs, along with AGI safety organization founders, and multiple AI unicorn founders."[240] Some of these guests wore shirts saying "Kurzweil was right" and "Does this help us get to safe AGI?"[241]

Hence, this form of pro-extinctionism is not a fringe view within the TESCREAL movement, or within Silicon Valley more generally. Jaron Lanier confirms this in reporting that "a lot" of the people in AI "believe that it would be good to wipe out people and that the AI future would be a better one." "Just the other day," he continues,

> I was at a lunch in Palo Alto and there were some young AI scientists there who were saying that they would never have a "bio baby" because as soon as you have a "bio baby," you get the "mind virus" of the [biological] world. And when you have the mind virus, you become committed to your human baby. But it's much more important to be committed to the AI of the future. And so to have human babies is fundamentally unethical.[242]

But this introduces a puzzle, since many of the loudest voices calling for efforts to *avert* human extinction arise from the TESCREAL community. How does this fit with their pro-extinctionist stance? There are two answers: the first is that while most people likely understand "human extinction" to mean *terminal* extinction, TESCREALists understand it to mean *final* extinction. Hence, they are advocating for us to avoid final rather than terminal extinction. Second, TESCREALists often define "human" or "humanity" in an idiosyncratic manner, as denoting both our species *and* whatever successors we might have, so long as they possess properties like consciousness, sentience, moral status, etc.[243] This more expansive definition implies that *Homo sapiens* could die out next year without "human extinction" having occurred. So long as we are replaced by successors with the right properties, then "humanity" will persist. Hence, when they talk about the survival of humanity, they do not mean the survival of *our species*; our survival matters *only insofar* as it is necessary to create posthuman successors. Confusingly, posthumanity would count as "humanity" on this expanded definition; that is, if posthumans possess the right properties, they would also be "humans."[244]

Some leading TESCREALists do not explicitly endorse the pro-extinctionism of digital eugenics. Yet most are nonetheless *indifferent* to our survival once posthumanity arrives. Torres calls this view *extinction neutralism*, citing Toby Ord as an example. Nowhere does Ord outright declare that *Homo sapiens* should vanish entirely, though he does write that "rising to our full potential for flourishing would likely involve us being transformed into something beyond the humanity of today."[245] In other words, he contends that humanity *must* create or become a new posthuman species.

Torres argues that extinction neutralism is likely indistinguishable from pro-extinctionism with respect to its practical consequences.[246] If we create a world that is ruled and run by posthumanity, why exactly would they keep us around? After all, as Shiller points out,

our resources are finite, and the same resources that might allow human beings to live—effort, land, energy, raw materials—could be more effectively spent on creating and sustaining artificial creatures. When that becomes the case, the beneficent thing to do is to choose that our children be artificial, rather than natural.[247]

After this, our species would fade away, thereby freeing up resources for our posthuman successors. These successors would have every reason to phase out humanity, and hence extinction neutralism seems to have pro-extinctionist implications *in practice*.

Torres thus argues that "nearly all TESCREALists fall somewhere on the spectrum between extinction neutralism and outright pro-extinctionism."[248] In both cases, the survival of our species in a posthuman "utopia" does not seem probable.

## 6. **Conclusion**

This article has outlined the TESCREAL ideologies, tracing their origins back to the 20th-century eugenics movement. We discussed the relationship between each ideology, and the reasons that Gebru and Torres defend the "strong" TESCREAL thesis. We then turned to some criticisms of the TESCREAL movement, which foregrounded questions about the feasibility of "safe" AGI, the dangerous of utopian-utilitarian reasoning, and the extent to which most of humanity, and our species itself, would survive in a posthuman world.

Finally, it is worth noting that many other scholars have converged upon similar ideas to those outlined by Gebru and Torres. Terms like "The Mindset" (from Douglas Rushkoff[249]), "technoeugenics" (from Anita Say Chan[250]), the "ideology of technological salvation" (from Adam Becker[251]), and "The Nerd Reich" (from Gil Duran[252]) all point to similar cultural phenomena: a broadly libertarian cluster of movements in the tech world that advocate a eugenic vision of cosmic utopia through the development of advanced technologies like AGI. The next step for scholars should be to unify and further develop this nascent literature, a task of some urgency given the resources being poured into the AGI race, amounting to an estimated $1.5 trillion dollars thus far[253]—money that could have been more wisely spent on tackling climate change, eliminating global poverty, and ensuring a positive future for all.

**Further Reading:**

Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636.

Alexander Thomas, *The Politics and Ethics of Transhumanism* (Bristol University Press, 2024). https://library.oapen.org/bitstream/handle/20.500.12657/91261/1/9781529239669.pdf.

Adam Becker, *More Everything Forever: AI Overlords, Space Empires, and Silicon Valley's Crusade to Control the Fate of Humanity* (Hachette, 2025).

Anita Say Chan, *Predatory Data: Eugenics in Big Tech and Our Fight for an Independent Future* (University of California Press, 2025).

Alice Crary, "The Toxic Ideology of Longtermism," *Radical Philosophy* 2.14 (April 2023). https://www.radicalphilosophy.com/commentary/the-toxic-ideology-of-longtermism.

Carol Adams, Alice Crary, and Lori Gruen, *The Good it Promises, the Harm it Does: Critical Essays on Effective Altruism* (Oxford University Press, 2023).

Emily Bender and Alex Hanna, *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want* (Harper 2025).

Karen Hao, *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI* (Penguin Publishing Group, 2025).

Douglas Rushkoff, *Survival of the Richest: Escape Fantasies of the Tech Billionaires* (W. W. Norton, 2022).

Elise Bohan, "A history of transhumanism," PhD diss., Doctoral dissertation, Macquarie University. http://hdl. handle. net/19 59.14/1271515 (2018).

Elise Bohan, *Future Superhuman: Our transhuman lives in a make-or-break century* (Newsouth, 2022).

1 Jean Gayon and Daniel Jacobi, eds., *L'éternal retour de l'eugénisme* (Paris: Presses Universitaires de France, 2006).

2 Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636.

3 See Susan B. Levin, *Posthuman Bliss? The Failed Promise of Transhumanism*. (Oxford University Press, 2021).

4 Nick Bostrom, "The Transhumanist FAQ," Version 2.1, *World Transhumanist Association*. https://nickbostrom.com/views/transhumanist.pdf.

5 Nick Bostrom, "Transhumanist values," In: Frederick Adams (ed.) *Ethical Issues for the 21st Century* (Philosophical Documentation Center Press, 2003).

6 Nick Bostrom, "Letter from Utopia," *Studies in Ethics, Law, and Technology* 2, no. 1 (2008/2020). https://doi.org/10.2202/1941-6008.1025.

7 Émile P. Torres, "The Acronym Behind Our Wildest AI Dreams and Nightmares," *Truthdig* (2023). https://www.truthdig.com/articles/the-acronym-behind-our-wildest-ai-dreams-and-nightmares/.

8 Michael A. Woodley of Menie, "Eugenics Movement," In: Zeigler-Hill, V., Shackelford, T.K. (eds) *Encyclopedia of Personality and Individual Differences* (Springer, 2020). https://doi.org/10.1007/978-3-319-24612-3_750

9 Quoted in Adam Rutherford, *Control: The Dark History and Troubling Present of Eugenics* (W. W. Norton, 2022).

10 Adam Rutherford, *Control: The Dark History and Troubling Present of Eugenics* (W. W. Norton, 2022).

11 Kathy L. Gaca, *The Making of Fornication: Eros, Ethics, and Political Reform in Greek Philosophy and Early Christianity* (University of California Press, 2017).

12 Mariska Leunissen, *From Natural Character to Moral Virtue in Aristotle* (Oxford University Press, 2017).

13 Richard S. Fogarty and Michael A. Osborne, "Eugenics in France and the Colonies," In: Alison Bashford and Philippa Levine (eds.) *The Oxford Handbook of the History of Eugenics* (Oxford, 2010).

14 Maria Sophia Quine, "The First-Wave Eugenic Revolution in Southern Europe: Science *sans frontièrs*," In: Alison Bashford and Philippa Levine (eds.) *The Oxford Handbook of the History of Eugenics* (Oxford, 2010).

15 Francis Galton, *Hereditary Genius*, Second Edition (MacMillan and Co., 1892).

16 See Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636.

17 Thanks to Anita Say Chan for improving this paragraph.

18 Anita Say Chan, *Predatory Data: Eugenics in Big Tech and Our Fight for an Independent Future* (University of California Press, 2025), and Peter Bowler, *Progress Unchained: Ideas of Evolution, Human History and the Future* (Cambridge University Press, 2021).

19 Alison Bashford and Philippa Levine, *The Oxford Handbook of the History of Eugenics* (Oxford University Press, 2010).

20 See Paul Weindling, "Julian Huxley and the continuity of eugenics in twentieth-century Britain," *Journal of Modern European History* 10, no. 4 (2012): 480-499.

21 Alison Bashford and Philippa Levine, "Introduction: Eugenics and the Modern World," In: Alison Bashford and Philippa Levine (eds.) *The Oxford Handbook of the History of Eugenics* (Oxford, 2010).

22 Maria Sophia Quine, "The First-Wave Eugenic Revolution in Southern Europe: Science *sans frontièrs*," In: Alison Bashford and Philippa Levine (eds.) *The Oxford Handbook of the History of Eugenics* (Oxford, 2010).

23 Nick Bostrom, "The Transhumanist FAQ," Version 2.1, *World Transhumanist Association*. https://nickbostrom.com/views/transhumanist.pdf, and Nick Bostrom, "Transhumanist values," In: Frederick Adams (ed.) *Ethical Issues for the 21st Century* (Philosophical Documentation Center Press, 2003).

24 Anders Sandberg. "Morphological freedom—Why we not just want it, but need it," In: Max More and Natasha Vita-More (eds.) *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future* (Wiley-Blackwell, 2013): 56-64.

25 Nick Bostrom, "Transhumanist values," In: Frederick Adams (ed.) *Ethical Issues for the 21st Century* (Philosophical Documentation Center Press, 2003).

26 Quoted in Émile P. Torres, "Nick Bostrom, Longtermism, and the Eternal Return of Eugenics," *Truthdig*. https://www.truthdig.com/articles/nick-bostrom-longtermism-and-the-eternal-return-of-eugenics-2/.

27 Julian Huxley, *Religion Without Revelation* (Harper & Brothers Publishers, 1927).

28 Julian Huxley, *New Bottles for New Wine* (Chatto & Windus, 1957).

29 Weindling P. 'Julian Huxley and the Continuity of Eugenics in Twentieth-century Britain'. J Mod Eur Hist. 2012 Nov 1;10(4):480-499. doi: 10.17104/1611-8944_2012_4. PMID: 25798079; PMCID: PMC4366572. See also Huxley, Julian. "Eugenics in evolutionary perspective." *Perspectives in Biology and Medicine* 6, no. 2 (1963): 155-187.

30 Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Viking, 2005).

31 Nick Bostrom, "Letter from Utopia," *Studies in Ethics, Law, and Technology* 2, no. 1 (2008/2020). https://doi.org/10.2202/1941-6008.1025.

32 Marco Vega and Peter Brietbart, PostHuman: An Introduction to Transhumanism (British Institute of Posthuman Studies, 2013). https://hpluspedia.org/wiki/PostHuman:_An_Introduction_to_Transhumanism.

33 Quoted in Elise Bohan, "A history of transhumanism," PhD diss., Doctoral dissertation, Macquarie University. http://hdl. handle. net/19 59.14/1271515, 2018.

34 Alexander Thomas, *The Politics and Ethics of Transhumanism* (Bristol University Press, 2024). https://library.oapen.org/bitstream/handle/20.500.12657/91261/1/9781529239669.pdf.

35 Max More, "The Extropian Principles," *Extropy* #6 (1990). https://fennetic.net/irc/extropy/ext6.pdf.

36 Elise Bohan, "A history of transhumanism," PhD diss., Doctoral dissertation, Macquarie University. http://hdl. handle. net/19 59.14/1271515 (2018).

37 Max More, "The Proactionary Principle," Vital Progress Summit I (2004). https://www.extropy.org/proactionaryprinciple.htm.

38 Max More, "The Proactionary Principle," *Extropic Thoughts* (2023). https://maxmore.substack.com/p/the-proactionary-principle.

39 "Extropy-chat—ExI chat list," Extropy.org (2025). https://lists.extropy.org/mailman/listinfo.cgi/extropy-chat.

40 Extropy Institute, Home page (2025). https://www.extropy.org/emaillists.htm.

41 Extropy Institute, "Events" page (2025). https://www.extropy.org/events.htm.

42 Alexander Thomas, *The Politics and Ethics of Transhumanism* (Bristol University Press, 2024). https://library.oapen.org/bitstream/handle/20.500.12657/91261/1/9781529239669.pdf.

43 As Hughes puts it, "most futurists actually lean left but the kind of dominant, hegemonic discourse is from people like Musk, Peter Thiel and people like that who are definitely not on the left." See James Hughes, "Moral Enhancement Technologies w/ James Hughes," Futures podcast (2022). https://futurespodcast.net/episodes/63-jameshughes.

44 Eliezer Yudkowsky, "The Singularitarian Principles v.1.0.2," Yudkowsky.net (2001). https://web.archive.org/web/20090107015106/http://yudkowsky.net/obsolete/principles.html.

45 David Thorstad, "Against the singularity hypothesis," *Philosophical Studies* (2024): 1-25. https://doi.org/10.1007/s11098-024-02143-5.

46 Eliezer Yudkowsky, "Three Major Singularity Schools," MIRI (2007). https://intelligence.org/2007/09/30/three-major-singularity-schools/.

47 Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Viking, 2005).

48 Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636

49 Admin, "MIRI: Artificial Intelligence: The Danger of Good Intentions," Future of Life Institute (2015). https://futureoflife.org/ai/ai-the-danger-of-good-intentions/.

50 Clint Rainey, "P(doom) is AI's latest apocalypse metric. Here's how to calculate your score," *Fast Company* (2023). https://www.fastcompany.com/90994526/pdoom-explained-how-to-calculate-your-score-on-ai-apocalypse-metric.

51 Elise Bohan, "A history of transhumanism," PhD diss., Doctoral dissertation, Macquarie University. http://hdl. handle. net/19 59.14/1271515, 2018.

52 Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Viking, 2005).

53 Eliezer Yudkowsky, "Re: SOCIETY: The Quiet Revolution," Extropians mailing list (1996). https://web.archive.org/web/20160414051408/https://extropians.weidai.com/extropians.96/4858.html.

54 Shane Legg, "The Transformative Potential of AGI—and When It Might Arrive," TED (2023). https://www.youtube.com/watch?v=kMUdrUP-QCs&t=175s.

55 Sam Altman, Twitter post (Jan. 4, 2025). https://x.com/sama/status/1875603249472139576?lang=en. And Sam Altman, "The Gentle Singularity," personal blog (2025). https://blog.samaltman.com/the-gentle-singularity.

56 Steven Levy, "If Anthropic Succeeds, a Nation of Benevolent AI Geniuses Could Be Born," *Wired* (2025). https://www.wired.com/story/anthropic-benevolent-artificial-intelligence/. And Shirin Ghaffary, "Anthropic Is Trying to Win the AI Race Without Losing Its Soul," *Bloomberg* (2025). https://archive.is/02ZdH.

57 Elon Musk, X post (2025). https://x.com/elonmusk/status/1893810875875889507.

58 Boris Groys, "Introduction: Russian Cosmism and the Technology of Immortality," *Russian Cosmism* (The MIT Press, 2018).

59 Boris Groys, "Introduction: Russian Cosmism and the Technology of Immortality," *Russian Cosmism* (The MIT Press, 2018).

60 Alexander Thomas, *The Politics and Ethics of Transhumanism* (Bristol University Press, 2024). https://library.oapen.org/bitstream/handle/20.500.12657/91261/1/9781529239669.pdf.

61 SingularityNET, LinkedIn (2025). https://www.linkedin.com/company/singularitynet/about/. SingularityNET, "About SingularityNET" (2025. https://singularitynet.io/aboutus/. And Ben Goertzel, *A Cosmist Manifesto: Practical Philosophy for the Posthuman Age* (Humanity+ Press, 2010).

62 Ben Goertzel, "I am Ben Goertzel, CEO of SingularityNET and TrueAGI. Ask Me Anything about AGI, the Technological Singularity, Robotics, the Future of Humanity, and Building Intelligent Machines!," *Reddit* (2024). https://www.reddit.com/r/Futurology/comments/1af02b8/comment/koi74uh/.

63 Ben Goertzel, *A Cosmist Manifesto: Practical Philosophy for the Posthuman Age* (Humanity+ Press, 2010).

64 Ben Goertzel, *A Cosmist Manifesto: Practical Philosophy for the Posthuman Age* (Humanity+ Press, 2010).

65 Ben Goertzel, *A Cosmist Manifesto: Practical Philosophy for the Posthuman Age* (Humanity+ Press, 2010).

66 Luke Muehlhauser, "Ben Goertzel on AGI as a Field," Machine Intelligence Research Institute (2013). https://intelligence.org/2013/10/18/ben-goertzel/.

67 radical_negative_one et al., "Ben Goertzel," LessWrong (2016). https://www.lesswrong.com/w/ben-goertzel#:~:text=kinds%20of%20explanation.-,Ben%20Goertzel%20is%20the%20Chairman%20at%20the%20AGI%20company%20Novamente,the%20idea,%20Coherent%20Aggregated%20Volition.

68 Ben Goertzel, "Who Coined the Term 'AGI'?," personal blog, goertzel.org (2011). https://goertzel.org/who-coined-the-term-agi/.

69 MIRI, "Singularity Summit: An Annual Conference on Science, Technology, and the Future," Machine Intelligence Research Institute (2025). https://goertzel.org/who-coined-the-term-agi/.

70 Ruby, Raemon, RobertM, and habryka, "Welcome to LessWrong!," LessWrong (2019). https://www.lesswrong.com/posts/bJ2haLkcGeLtTWaD5/welcome-to-lesswrong. Note that the wording here is similar to that of Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636.

71 Nick Bostrom, "Letter from Utopia," *Studies in Ethics, Law, and Technology* 2, no. 1 (2008/2020). https://doi.org/10.2202/1941-6008.1025.

72 Eliezer Yudkowsky, X post (2023). https://x.com/ESYudkowsky/status/1654171832994365475.

73 The standard story told about EA is that it started out being focused on global poverty, and only later pivoted toward longtermist cause areas. However, Mollie Gleiberman convincingly shows that this is not the case: EA was longtermism-focused from its very origins. See Mollie Gleiberman, "Effective Altruism and the Strategic Ambiguity of 'Doing Good,'" IOB Discussion Papers from Universiteit Antwerpen, Institute of Development Policy (2023). https://medialibrary.uantwerpen.be/files/8518/61565cb6-e056-4e35-bd2e-d14d58e35231.pdf. And Mollie Gleiberman, "Effective Altruism: Doing Transhumanism Better," IOB Discussion Papers from Universiteit Antwerpen, Institute of Development Policy (2023). https://medialibrary.uantwerpen.be/files/8518/c8d22373-5690-4a9b-a472-aa8bde961caf.pdf.

74 William MacAskill, *Doing Good Better: Effective Altruism and How You Can Make a Difference* (Random House, 2015).

75 Anthony Kalulu, "Effective Altruism Is Worse than Traditional Philanthropy in the Way It Excludes the Extreme Poor in the Global South," Dear Humanity (2022). https://dear-humanity.org/effective-altruism-worse-for-poor/.

76 William MacAskill, *Doing Good Better: Effective Altruism and How You Can Make a Difference* (Random House, 2015).

77 William MacAskill, "Replaceability, career choice, and making a difference," *Ethical Theory and Moral Practice* 17 (2014): 269-283.

78 David Yaffe-Bellany, "A Crypto Emperor's Vision: No Pants, His Rules," *New York Times* (2022). https://archive.is/InTe7.

79 Adam Fisher, "Sam Bankman-Fried Has a Savior Complex—And Maybe You Should Too," *Sequoia* (2022). https://archive.ph/xy4MR#selection-163.0-168.0.

80 Richard Chappell, Darius Meissner, and William MacAskill, *An Introduction to Utilitarianism: From Theory to Practice* (Hackett Publishing Company, 2024).

81 Carol Adams, Alice Crary, and Lori Gruen, *The Good It Promises, the Harm It Does* (Oxford University Press, 2023).

82 Émile P. Torres, "Fraud, Lies, Exploitation and Eugenic Fantasies," *Truthdig* (2023). https://www.truthdig.com/articles/effective-altruism-is-a-welter-of-fraud-lies-exploitation-and-eugenic-fantasies/.

83 Eliezer Yudkowsky, "Torture vs. Dust Specks," LessWrong (2007). https://www.lesswrong.com/posts/3wYTFWY3LKQCnAptN/torture-vs-dust-specks.

84 Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984).

85 Stéphane Zuber, Nikhil Venkatesh, Torbjörn Tännsjö, Christian Tarsney, H. Orri Stefánsson, Katie Steele, Dean Spears et al., "What should we agree on about the repugnant conclusion?," *Utilitas* 33, no. 4 (2021): 379-383. And William MacAskill, *What We Owe the Future* (Basic Books, 2022).

86 William MacAskill, *What We Owe the Future* (Basic Books, 2022). See also Hilary Greaves and William MacAskill, "The Case for Strong Longtermism," Global Priorities Institute, GPI Working Paper No. 5-2021 (2021). https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf.

87 Carl Sagan, "Nuclear war and climatic catastrophe: Some policy implications," *Foreign Affairs* 62, no. 2 (1983): 257-292.

88 Toby Newberry, "How Many Lives Does the Future Hold?," Global Priorities Institute, GPI Technical Report No. T2-2021 (2021).

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

90 Nick Beckstead, *On the overwhelming importance of shaping the far future*. Rutgers The State University of New Jersey, School of Graduate Studies (2013). https://www.proquest.com/docview/1442191960?fromopenview=true&pq-origsite=gscholar&sourcetype=Dissertations%20&%20Theses.

91 Benjamin Todd, "Future Genrations and Their Moral Significance," 80,000 Hours (2017). https://web.archive.org/web/20210311195203/https://80000hours.org/articles/future-generations/.

92 William MacAskill, *What We Owe the Future* (Basic Books, 2022).

93 For an account of (roughly speaking) conditional caring about future people, see Johann Frick, "Conditional reasons and the procreation asymmetry," *Philosophical Perspectives* 34, no. 1 (2020): 53-87.

94 William MacAskill, *What We Owe the Future* (Basic Books, 2022), italics added.

95 Toby Newberry, "How Many Lives Does the Future Hold?," Global Priorities Institute, GPI Technical Report No. T2-2021 (2021). Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, no. 3 (2003). And Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014).

96 Émile P. Torres, "Colonization, Consciousness, and Longtermism," In: Mirko Daniel Garasic and Marcello Di Paola (eds.) *The Philosophy of Outer Space: Explorations, Controversies, Speculations* (Routledge, 2024).

97 Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Jounral of Evolution and Technology* 9, no. 1 (2002).

98 Matjaz Gams and Jana Krivec, "Demographic Analysis of Fertility Using Data Mining Tools," *Informatica* 32, no. 2 (2008).

99 Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, no. 3 (2003).

[100] Nick Beckstead, *On the overwhelming importance of shaping the far future*. Rutgers The State University of New Jersey, School of Graduate Studies (2013). For discussion, see Alice Crary, "The Toxic Ideology of Longtermism," *Radical Philosophy* 2, no.14 (2023).

[101] Émile P. Torres, "Against Longtermism," *Aeon* (2021). https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo.

[102] Nick Bostrom, "The Future of Humanity," In: Jan-Kyrre Berg Olsen, Evan Selinger, and Soren Riis (eds.) *New Waves in Philosophy of Technology* (Palgrave McMillan, 2009).

[103] Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Jounral of Evolution and Technology* 9, no. 1 (2002).

[104] Nick Bostrom, "Transhumanist values," In: Frederick Adams (ed.) *Ethical Issues for the 21st Century* (Philosophical Documentation Center Press, 2003).

[105] Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, no. 3 (2003).

[106] Nick Bostrom and Toby Ord, "The Reversal Test: Eliminating Status Quo Bias in Applied Ethics," *Ethics* 116, no. 4 (2006). And Toby Ord, CV, personal website (2025). https://www.amirrorclear.net/files/toby-ord-cv.pdf.

[107] Nick Beckstead, LessWrong (2025). https://www.lesswrong.com/users/nick_beckstead?from=search_page.

[108] See Elise Bohan, "Transhumanism in EA and Beyond," EAGxAustralia (2022). https://www.youtube.com/watch?v=jKvfa4-wMdo&t=1235s.

[109] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014).

[110] Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Viking, 2005).

[111] Zoe Cremer, "Objections to Value-Alignment between Effective Altruists," Effective Altruism Forum (2020). https://forum.effectivealtruism.org/posts/DxfpGi9hwvwLCf5iQ/objections-to-value-alignment-between-effective-altruists.

[112] OpenPhil, "Our Team, Dustin Moskovitz," Open Philanthropy (2025). https://www.openphilanthropy.org/about/team/dustin-moskovitz/.

[113] Results obtained from www.openbook.fyi on June 1, 2025.

[114] Émile P. Torres, "The Acronym Behind Our Wildest AI Dreams and Nightmares," *Truthdig* (2023). https://www.truthdig.com/articles/the-acronym-behind-our-wildest-ai-dreams-and-nightmares/.

[115] Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636. And Eliezer Yudkowsky, Autobiography, personal website (2000). https://web.archive.org/web/20010205221413/http://sysopmind.com/eliezer.html#timeline_birth.

[116] Ruha Benjamin, "The New Artificial Intelligentsia," *Los Angeles Review of Books* (2024). https://lareviewofbooks.org/article/the-new-artificial-intelligentsia/.

[117] Carla Cremer, "How Effective Altruists Ignored Risk," *Vox* (2023). https://www.vox.com/future-perfect/23569519/effective-altrusim-sam-bankman-fried-will-macaskill-ea-risk-decentralization-philanthropy.

[118] Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636. See also Ruha Benjamin, "The New Artificial Intelligentsia," *Los Angeles Review of Books* (2024). https://lareviewofbooks.org/article/the-new-artificial-intelligentsia/.

[119] See Alexander Thomas, *The Politics and Ethics of Transhumanism* (Bristol University Press, 2024). https://library.oapen.org/bitstream/handle/20.500.12657/91261/1/9781529239669.pdf. The TESCREAL movement also has interesting connections to the psychedelics industry, a topic that I do not have space to discuss here. See Neşe Devenot, "TESCREAL Hallucinations: Psychedelic and AI Hype as Inequality Engines," *Journal of Psychedelic Studies* 7, no. S1 (2023): 22-39.

[120] The view of this author is that not everyone in the TESCREAL community *overtly* holds problematic—e.g., racist, sexist, ableist, etc.—views. Some members of the movement likely do; others are perhaps best characterized by Audra Mitchell and Aadita Chaudhury in their article about how contemporary discourse on "the end of the world" often conflates this with "the end of whiteness." They write:

> Whiteness is remarkable in its ability to render itself invisible to those who possess and benefit
>
> from it. Many, if not most, of the … authors of 'end of the world' discourses seem unaware of its integral influence on their thinking, and would almost certainly be horrified at the thought of their work entrenching racialized injustices. We are not suggesting that these authors espouse explicit, intentional and/or extreme racist ideals, on which much public discussion by white people of racism tends to focus. … On the contrary, we work to center broad, everyday, structural ways in which underlying *logics* of whiteness and white supremacy frame and permeate *mainstream* paradigms and discourses, including those identified as liberal, humanitarian, or progressive. Even amongst white people who consciously and explicitly disavow racism, unconscious, habitual, normalized, structurally-embedded assumptions circulate, and are reproduced in ways that perpetuate race as a global power structure.

See Audra Mitchell and Aadita Chaudhury, "Worlding Beyond 'the' 'end' of 'the world': White Apocalyptic Visions and BIPOC Futurisms," *International Relations* 34, no. 3 (2020): 309-332.

[121] Shane Legg, "The Transformative Potential of AGI—and When It Might Arrive," TED (2023). https://www.youtube.com/watch?v=kMUdrUP-QCs&t=175s.

[122] Eliezer Yudkowsky, "The Magnitude of His Own Folly," LessWrong (2008). https://www.lesswrong.com/posts/fLRPeXihRaiRo5dyX/the-magnitude-of-his-own-folly?commentId=4CvwN95eQev9DLtpF. And SL4, "Welcome to SL4!" (2015). http://sl4.org/#:~:text=The%20SL4%20mailing%20list%20is,strategies%20to%20accelerate%20the%20Singularity.

[123] Keach Hagey, "How Peter Thiel's Relationship with Eliezer Yudkowsky Launched the AI Revolution," *Wired* (2025). https://archive.is/c0ToC#selection-2645.318-2645.332.

[124] Shane Legg, "Measuring Machine Intelligence," Singularity Summit 2010 (2010). https://www.youtube.com/watch?v=0ghzG14dT-w&t=22s.

[125] DeepAI, "Shane Legg," DeepAI (2025). https://deepai.org/profile/shane-legg. And Demis Hassabis, "A Systems Neuroscience Approach to Building AGI," Singularity Summit 2010 (2010). https://www.youtube.com/watch?v=Qgd3OK5DZWI&t=1s.

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

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

[128] CFI, "Jaan Tallinn," Leverhulme Centre for the Future of Intelligence (2025). https://www.lcfi.ac.uk/people/jaan-tallinn.

[129] Alex Herd, "DeepMind announces ethics group to focus on problems of AI," *The Guardian* (2017). https://www.theguardian.com/technology/2017/oct/04/google-deepmind-ai-artificial-intelligence-ethics-group-problems.

[130] Rob Wiblin and Keiran Harris, "Mustafa Suleyman on Getting Washington and Silicon Valley to Tame AI," 80,000 Hours (2023). https://80000hours.org/podcast/episodes/mustafa-suleyman-getting-washington-and-silicon-valley-to-tame-ai/.

[131] OpenAI, "Introducing OpenAI," OpenAI website (2015). https://openai.com/index/introducing-openai/.

[132] Read Albergotti, "The Co-Founder of Skype Invested in Some of AI's Hottest Startups—but He Thinks He Failed," *Semafor* (2023). https://www.semafor.com/article/04/28/2023/co-founder-of-skype-invested-in-hot-ai-startups-but-thinks-he-failed.

[133] Marc Andreessen, Twitter profile (2023). https://web.archive.org/web/20230523005947/https://twitter.com/pmarca.

[134] OpenPhil, "OpenAI—General Support," Open Philanthropy (2017). https://www.openphilanthropy.org/grants/openai-general-support/.

[135] Cade Metz, "The ChatGPT King Isn't Worried, but He Knows You Might Be," *New York Times* (2023). https://archive.is/SMSea.

[136] Elizabeth Weil, "Sam Altman Is the Oppenheimer of Our Age," *Intelligencer* (2023). https://nymag.com/intelligencer/article/sam-altman-artificial-intelligence-openai-profile.html.

[137] Sam Altman, Twitter post (2023). https://x.com/sama/status/1621621724507938816?lang=en; lightly edited for clarity.

[138] Sam Altman, "Machine Intelligence, part 1," personal blog (2015). https://blog.samaltman.com/machine-intelligence-part-1.

[139] Sam Altman, Twitter post (2025). https://x.com/sama/status/1875603249472139576?lang=en; lightly edited for clarity.

[140] Antonio Regalado, "A Startup Is Pitching a Mind-Uploading Service that Is '100 Percent Fatal,'" *MIT Technology Review* (2018).

[141] Sam Altman, Twitter post (2022). https://x.com/sama/status/1559011065899282432?lang=ga.

[142] Sam Altman, Twitter post (2022). https://x.com/sama/status/1540782004836241409.

[143] Victor Tangermann, "Sam Altman Seems to Imply that OpenAI Is Building God," *Futurism* (2023). https://futurism.com/sam-altman-imply-openai-building-god. And Connie Loizos, "StrictlyVC in Conversation with Sam Altman, part two (OpenAI)," YouTube video. https://www.youtube.com/watch?v=ebjkD1Om4uw&t=1339s.

[144] Quoted in Émile P. Torres, "The Madness of the Race to Build Artificial General Intelligence," *Truthdig* (2024). https://www.truthdig.com/articles/the-madness-of-the-race-to-build-artificial-general-intelligence/.

[145] I. J. Good, "Speculations concerning the first ultraintelligent machine," *Advances in Computers*, 6 (1956), 31-88.

[146] I. J. Good, "Speculations on perceptrons and other automata," *International Business Machines Corporation* (1959).

[147] Joe Rogan, "Elon Musk," Joe Rogan Experience (2025). https://www.youtube.com/watch?v=sSOxPJD-VNo&t=8414s.

[148] Big Think, "The Intelligence Explosion: Nick Bostrom on the Future of AI," Big Think (2023). https://www.youtube.com/watch?v=1WcpN4ds0iY.

[149] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014).

[150] Big Think, "Nick Bostrom on the Birth of Superintelligence," Big Think (2025). https://bigthink.com/series/the-big-think-interview/superintelligence/.

[151] Karen Hao, *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI* (Penguin Publishing Group, 2025).

[152] Quoted in Karen Hao, *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI* (Penguin Publishing Group, 2025); italics added.

[153] LessWrong, "Dario_Amodei," LessWrong (2025). https://www.lesswrong.com/users/dario_amodei?from=search_page.

[154] Shirin Ghaffary, "Anthropic Is Trying to Win the AI Race Without Losing Its Soul," *Bloomberg* (2025). https://archive.is/02ZdH.

[155] Kevin Roose, "Inside the White-Hot Center of AI Doomerism," *New York Times* (2023). https://archive.is/pt5fQ.

[156] Karen Hao, *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI* (Penguin Publishing Group, 2025).

[157] Anthropic, "Anthropic Raises $124 million To Build More Reliable, General AI Systems," Anthropic website (2021). https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems. The amount that Tallinn donated was provided through personal communication with Tallinn.

[158] Jeremy Kahn, "Why Sam Bankman-Fried's FTX Debacle is Roiling AI Research," *Fortune* (2022). https://fortune.com/2022/11/15/sam-bankman-fried-ftx-collapse-a-i-safety-research-effective-altruism-debacle/.

[159] NVIDIA, "NVIDIA Ethernet Networking Accelerates World's Largest AI Supercomputer, Built by xAI," NVIDIA website (2024). https://nvidianews.nvidia.com/news/spectrum-x-ethernet-networking-xai-colossus.

[160] Anthony Cuthbertson, "Value Is Building Brain-Computer Interface for Fully-Immersive Video Games, President Reveals," *The Independent* (2021). https://www.the-independent.com/tech/valve-brain-computer-interface-video-game-b1792225.html.

[161] Elon Musk, X post (2025). https://x.com/elonmusk/status/1893810875875889507.

162 Elon Musk, Twitter post, (2022). https://x.com/elonmusk/status/1554335028313718784?lang=en.

163 Wachtell, Lipton, Rosen & Katz, "Exhibit A," in *Twitter, Inc. v Musk, et al.* (2022). https://s3.documentcloud.org/documents/23108357/redacted-version-of-exhibits-a-j-to-letter-to-the-honorable-kathaleen-st-j-mccormick-from-edward-b.pdf?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosam&stream=top.

164 David Martin Davies, "Elon Musk, Twitter and the Future of Human Civilization," Texas Public Radio (2023). https://www.tpr.org/podcast/the-source/2023-10-09/elon-musk-twitter-and-the-future-of-human-civilization. And Elon Musk, Twitter post (2022). https://x.com/elonmusk/status/1597405399040217088?lang=en.

165 Liv Boeree, Twitter post (2022). https://x.com/Liv_Boeree/status/1529158437585752064.

166 Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, no. 3 (2003).

167 Quoted in Ross Andersen, "Exodus," *Aeon* (2014). https://aeon.co/essays/elon-musk-puts-his-case-for-a-multi-planet-civilisation.

168 NYTE, "Elon Musk on Advertisers, Trust and the 'Wild Storm' in His Mind," New York Times Events (2023). https://www.youtube.com/watch?v=2BfMuHDfGJI&t=3439s.

169 Elon Musk, Twitter post (2014). https://x.com/elonmusk/status/495759307346952192.

170 Émile P. Torres, "Elon Musk, Twitter and the Future: His Long-Term Vision Is even Weirder than You Think," *Salon* (2022). https://www.salon.com/2022/04/30/elon-musk-twitter-and-the-future-his-long-term-vision-is-even-weirder-than-you-think/.

171 Jelor Gallego, "Elon Musk Is Looking to Kickstart Transhuman Evolution with 'Brain Hacking' Tech," *Futurism* (2016). https://futurism.com/elon-musk-is-looking-to-kickstart-transhuman-evolution-with-brain-hacking-tech.

172 Elise Bohan, *Future Superhuman: Our transhuman lives in a make-or-break century* (Newsouth, 2022).

173 SpaceX, Twitter post (2013). https://x.com/spacex/status/380084890907074560.

174 Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636

175 TW123 and Dan Hendrycks, "Introduction to Pragmatic AI Safety," Effective Altruism Forum (2022). https://forum.effectivealtruism.org/posts/MskKEsj8nWREoMjQK/introduction-to-pragmatic-ai-safety-pragmatic-ai-safety-1.

176 Émile P. Torres, "'Effective Accelerationism' and the Pursuit of Cosmic Utopia," *Truthdig* (2023). https://www.truthdig.com/articles/effective-accelerationism-and-the-pursuit-of-cosmic-utopia/.

177 This echoes a claim from Wayne Lapierre, the former CEO of the National Rifle Association in the United States, that "the only thing that stops a bad guy with a gun, is a good guy with a gun." See Peter Overby, "NRA: 'Only Thing that Stops a Bad Guy with a Gun Is a Good Guy with a Gun," NPR (2012). https://www.npr.org/2012/12/21/167824766/nra-only-thing-that-stops-a-bad-guy-with-a-gun-is-a-good-guy-with-a-gun.

178 Marc Andreessen, "The Techno-Optimist Manifesto," Andreessen-Horowitz (2023). https://a16z.com/the-techno-optimist-manifesto/.

179 Eliezer Yudkowsky, "Pausing AI Developments Isn't Enough. We Need to Shut it All Down," *Time* (2023). https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/, italics added.

180 Gretta Duleba, "MIRI 2024 Communications Strategy," Machine Intelligence Research Institute (2024). https://intelligence.org/2024/05/29/miri-2024-communications-strategy/.

181 Rob Bensinger, X post (2023). https://x.com/robbensinger/status/1727387469174407551.

182 Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636. See also Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014), ch. 2.

183 Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636.

184 Heidy Khlaaf, "Toward Comprehensive Risk Assessments and Assurance of AI-Based Systems," Trail of Bits (2023). https://www.trailofbits.com/documents/Toward_comprehensive_risk_assessments.pdf.

185 Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636.

186 Sam Altman, "Planning for AGI and Beyond," OpenAI website (2023). https://openai.com/index/planning-for-agi-and-beyond/.

187 OpenAI Charter, OpenAI website (2025). https://openai.com/charter/.

188 Dave Bergmann and Cole Stryker, "What Is Artificial General Intelligence (AGI)?," IBM website (2024). https://www.ibm.com/think/topics/artificial-general-intelligence.

189 Anca Dragan, Rohin Shah, Four Flynn, and Shane Legg, "Taking a Responsible Path to AGI," DeepMind website (2025). https://deepmind.google/discover/blog/taking-a-responsible-path-to-agi/.

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

191 Dario Amodei, "Machines of Loving Grace," personal website (2024). https://www.darioamodei.com/essay/machines-of-loving-grace.

192 Timnit Gebru and Émile P. Torres, "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence," *First Monday* 4, no. 1 (2024). https://doi.org/10.5210/fm.v29i4.13636.

193 Big Think, "Nick Bostrom on the Birth of Superintelligence," Big Think (2025). https://bigthink.com/series/the-big-think-interview/superintelligence/.

194 Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Jounral of Evolution and Technology* 9, no. 1 (2002).

195 Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, no. 3 (2003).

196 Nick Bostrom, *Deep Utopia: Life and Meaning in a Solved World* (IdeaPress Publishing, 2024).

197 See Émile P. Torres, "Longtermism Poses a Real Threat to Humanity," *The New Statesman* (2023). https://www.newstatesman.com/ideas/2023/08/longtermism-threat-humanity.

198 Billy Perrigo, "Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic," *Time* (2023). https://time.com/6247678/openai-chatgpt-kenya-workers/. See also Karen Hao, *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI* (Penguin Publishing Group, 2025).

199 Émile P. Torres, "On the Extinction of Humanity," *Synthese* (forthcoming). And Émile P. Torres, "If Artificial Superintelligence Were to Cause Our Extinction, Would That Be So Bad?," *Canadian Journal of Bioethics* (forthcoming).

200 See Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, no. 3 (2003).

201 Nick Bostrom, "Existential Risk Prevention as Global Priority," *Global Policy* 4, no. 1 (2013): 15-31.

202 Quoted in Adam Becker, *More Everything Forever AI Overlords, Space Empires, and Silicon Valley's Crusade to Control the Fate of Humanity* (Basic Books, 2025).

203 Eliezer Yudkowsky, X post (2023). https://x.com/ESYudkowsky/status/1726284483748659691.

204 Adam Becker, *More Everything Forever AI Overlords, Space Empires, and Silicon Valley's Crusade to Control the Fate of Humanity* (Basic Books, 2025).

205 Nahlah Ayed, "Techno-Utopia or the Billionaires' Wet Dream," CBC *Ideas* (2025). https://metacast.app/podcast/ideas/JJWBzjpI/techno-utopia-or-the-billionaires-wet-dream/vqAXzxTS.

206 Hilary Greaves and William MacAskill, "The Case for Strong Longtermism," Global Priorities Institute, GPI Working Paper No. 5-2021 (2021). https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf.

207 Quoting Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (Bloomsbury Publishing, 2020).

208 Émile P. Torres, "Against Longtermism," *Aeon* (2021). https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo.

209 Nick Bostrom, "Existential Risk Prevention as Global Priority," *Global Policy* 4, no. 1 (2013): 15-31.

210 Olle Häggström, *Here Be Dragons: Science, Technology and the Future of Humanity* (Oxford University Press, 2016).

211 Note that Häggström is otherwise sympathetic with longtermism.

212 See Helga Kuhse and Peter Singer, Should the Baby Live?The Problem of Handicapped Infants (Oxford University Press, 1985). And Peter Singer, Nick Beckstead, and Matt Wage, "Preventing Human Extinction," Effective Altruism Forum (2013). https://forum.effectivealtruism.org/posts/tXoE6wrEQv7GoDivb/preventing-human-extinction.

213 Peter Singer, "The Hinge of History," *Project Syndicate* (2021). https://www.project-syndicate.org/commentary/ethical-implications-of-focusing-on-extinction-risk-by-peter-singer-2021-10.

214 Quoted in Émile P. Torres, "Before It's Too Late, Buddy," *Truthdig* (2023). https://www.truthdig.com/articles/before-its-too-late-buddy/.

215 OpenPhil, "Century Fellowship—2022 Cohort," Open Philanthropy (2022). https://www.openphilanthropy.org/grants/century-fellowship-2022-cohort/.

216 Akash Wasil, "Experience," LinkedIn (2025). https://www.linkedin.com/in/akash-wasil/details/experience/.

217 Thomas Larsen, Profile, LinkedIn (2025). https://www.linkedin.com/in/thomas-larsen/.

218 See Émile P. Torres, "Before It's Too Late, Buddy," *Truthdig* (2023). https://www.truthdig.com/articles/before-its-too-late-buddy/.

219 Lili Xia, Alan Robock, Kim Scherrer, et al., "Global Food Insecurity and Famine from Reduced Crop, Marine Fishery and Livestock Production Due to Climate Disruption from Nuclear War Soot Injection," *Nature Food* 3 (2022): 586-596.

220 Eliezer Yudkowsky, "Pausing AI Developments Isn't Enough. We Need to Shut it All Down," *Time* (2023). https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/.

221 Quoted in Émile P. Torres, "The Acronym Behind Our Wildest AI Dreams and Nightmares," *Truthdig* (2023). https://www.truthdig.com/articles/the-acronym-behind-our-wildest-ai-dreams-and-nightmares/.

222 Roman Yampolskiy, "Minimum Viable Human Population with Intelligent Interventions," ResearchGate (2018). DOI:10.13140/RG.2.2.33337.83042.

223 Émile P. Torres, "What the Sam Bankman-Fried Debacle Can Teach Us About 'Longtermism,'" *Salon* (2022). https://www.salon.com/2022/11/20/what-the-sam-bankman-fried-debacle-can-teach-us-about-longtermism/.

224 Émile P. Torres, "Does AGI Really Threaten the Survival of the Species?," *Truthdig* (2023). https://www.truthdig.com/articles/does-agi-really-threaten-the-survival-of-the-species/. And Émile P. Torres, "The Future of the Future of Humanity Institute," *Substack* (2024). https://xriskology.substack.com/p/the-future-of-the-future-of-humanity.

225 Quoted in Émile P. Torres, "Dangerous Visions: How the Quest for Utopia Could Lead to Catastrophe," *Salon* (2023). https://www.salon.com/2023/07/29/visions-how-the-quest-for-utopia-could-lead-to/. Bielskyte further argues for an alternative to utopia and dystopia, which she calls "protopia." This is intended to fall *not* within the utopia-dystopia framework, but entirely outside of it—a fundamentally different way of thinking about normative futurological questions. I am very sympathetic to Bielskyte's protopianism, as well as other alternatives like "throughtopia," which Rupert Read describes as being "about getting through what is coming responsibly, transformatively in the best way we can." See, respectively, Monika Bielskyte, "Protopia Futures," *Medium* (2021). https://medium.com/protopia-futures/protopia-futures-framework-f3c2a5d09a1e, and Rupert Read, "Why Apocalypses Aren't Enough: Imagining the New Political Culture," The Climate Majority Project (2023). https://rupertread.net/audio-video/2023/why-apocalypses-arent-enough-imagining-the-new-political-culture-the-green-gathering/.

226 Robert Sparrow, "A not-so-new eugenics: Harris and Savulescu on human enhancement," *The Hastings Center Report* 41, no. 1 (2011): 32-42.

227 William MacAskill, *What We Owe the Future* (Basic Books, 2022).

228 WWF, "Living Planet Report: Devastating 69% Drop in Wildlife Populations since 1970," World Wildlife Fund (2022). https://www.wwf.eu/?7780966/WWF-Living-Planet-Report-Devastating-69-drop-in-wildlife-populations-since-1970.

229 Émile P. Torres, "Dangerous Visions: How the Quest for Utopia Could Lead to Catastrophe," *Salon* (2023). https://www.salon.com/2023/07/29/visions-how-the-quest-for-utopia-could-lead-to/.

230 Sam Altman actually recognizes this in saying that "if I weren't in on this [the AGI race] I'd be, like, why do these fuckers get to decide what happens to me?" He thus reported that OpenAI is "planning a way to allow wide swaths of the world to elect representatives to a new governance board." Yet, as Gary Marcus pointed out, OpenAI has done no such thing—it has, indeed, established a board consisting most mostly men from the Global North, including Altman himself. See Gary Marcus, "Sam Altman, Then and Now," *Substack* (2024). https://garymarcus.substack.com/p/sam-altman-then-and-now.

231 Émile P. Torres, "A Tale of Two Extinctions," *Truthdig* (2025). https://www.truthdig.com/articles/a-tale-of-two-extinctions/.

232 For use of the term "digital eugenics," see Max Tegmark, "Max Tegmark Warns about 'Digital Eugenics' in AI," Nonzero (2025). https://www.youtube.com/shorts/jCpmdb_0Cf0.

233 Hans Moravec, "Human Culture: A Genetic Takeover Underway," In: Christopher Langton (ed.) *Artificial Life Proceedings Of An Interdisciplinary Workshop On The Synthesis And Simulation Of Living Systems* (Taylor & Francis 2010).

234 Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Knopf Doubleday Publishing Group, 2017).

235 Derek Shiller, "In defense of artificial replacement," *Bioethics* 31, no. 5 (2017): 393-399.

236 Quoted in Émile P. Torres, "The Endgame of Edgelord Eschatology," *Truthdig* (2025). https://www.truthdig.com/articles/the-endgame-of-edgelord-eschatology/.

237 Daniel Faggella, "Eliezer Yudkowsky: Human Augmentation as a Safer AGI Pathway," *The Trajectory* (2025). https://www.youtube.com/watch?v=YlsvQO0zDiE&t=3899s.

238 Émile P. Torres, "The Endgame of Edgelord Eschatology," *Truthdig* (2025). https://www.truthdig.com/articles/the-endgame-of-edgelord-eschatology/.

239 See Émile P. Torres, X post (2025). https://x.com/xriskology/status/1929707364728688747.

240 Daniel Faggella, email listserv. *The Trajectory* (2025). https://ao8p8.r.a.d.sendibm1.com/mk/mr/sh/6rqJ8GoudeITQRjoZgjbRkcY59a/WqU3OZDBd9MG.

241 Kylie Robinson, "Inside the AI Party at the End of the World," *Wired* (2025). https://archive.is/EHcAK.

242 Sean Illing, "Will AI Become God? That's the Wrong Question," *Vox* (2025). https://www.vox.com/the-gray-area/407154/jaron-lanier-ai-religion-progress-criticism.

243 For examples, see Émile P. Torres, "On the Extinction of Humanity," *Synthese* (forthcoming).

244 Émile P. Torres, "On the Extinction of Humanity," *Synthese* (forthcoming). And Émile P. Torres, "If Artificial Superintelligence Were to Cause Our Extinction, Would That Be So Bad?," *Canadian Journal of Bioethics* (forthcoming).

245 Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (Bloomsbury Publishing, 2020).

246 Émile P. Torres, "On the Extinction of Humanity," *Synthese* (forthcoming). Émile P. Torres, "If Artificial Superintelligence Were to Cause Our Extinction, Would That Be So Bad?," *Canadian Journal of Bioethics* (forthcoming). Émile P. Torres, "Should Humanity Go Extinct? Examining the Arguments for Pro-Extinctionism" (under review).

247 Derek Shiller, "In defense of artificial replacement," *Bioethics* 31, no. 5 (2017): 393-399.

248 Émile P. Torres, "A Tale of Two Extinctions," *Truthdig* (2025). https://www.truthdig.com/articles/a-tale-of-two-extinctions/.

249 Douglas Rushkoff, *Survival of the Richest: Escape Fantasies of the Tech Billionaires* (W. W. Norton, 2022).

250 Anita Say Chan, *Predatory Data: Eugenics in Big Tech and Our Fight for an Independent Future* (University of California Press, 2025).

251 Adam Becker, *More Everything Forever AI Overlords, Space Empires, and Silicon Valley's Crusade to Control the Fate of Humanity* (Basic Books, 2025).

252 See Gil Duran's blog, "The Nerd Reich": https://www.thenerdreich.com/.

253 Lauren Smart and Sam Hsu, "The AI-Energy Nexus Will Determine AI's Impact. We Must Account for it Better." *World Economic Forum* (2025). https://www.weforum.org/stories/2025/12/ai-energy-nexus-ai-future/.