The Myth of AI, Existential Threat, Why The Myth Persists, and What is to be Done About It

Robert Hanna

(RAND, 2023)

1. Introduction

No digital computing systems or digital technology, no matter how sophisticated or tricked-out with high-tech bells and whistles, will ever be able to equal or exceed the essentially embodied innate mental capacities or powers of rational human animals[1]—including, of course, the readers of this very sentence. Not even *in principle* can such systems or technology match or surpass our capacities or powers. To be sure, digital computing systems or digital technology can carry out certain operations *much faster* and

[1] Or of rational *non*-human animals, if there are any. For the record, the basic essentially embodied innate mental capacities or powers are: (i) *consciousness*, i.e., subjective experience, (ii) *self-consciousness*, i.e., consciousness of one's own consciousness, second-order consciousness, (iii) *caring*, i.e., desiring, emoting, or feeling, (iv) *sensible cognition*, i.e., sense-perception, memory, or imagination, (v) *intellectual cognition*, i.e., conceptualizing, believing, judging, or inferring, (vi) *volition*, i.e., deciding, choosing, or willing, and (vii) *free agency*, i.e., free will and practical agency. In human animals, the unified set of these capacities constitutes our *rational human mindedness*, which is the same as our *human real personhood* (Hanna, 2018). See also section 6 below.

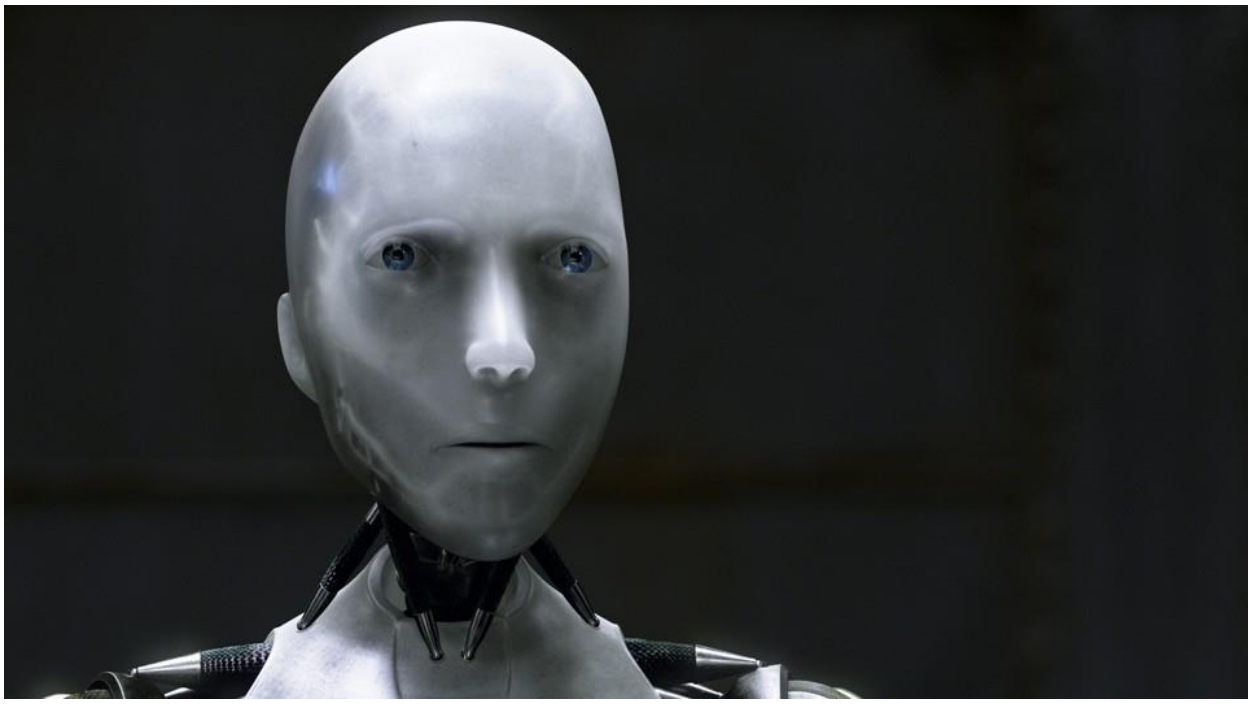
more accurately than we can. But that's not a fact about *the nature, scope, and limits of our mental capacities or powers*, but rather only a fact about *the applications of those powers to certain mechanical tasks*, and no more philosophically exciting or significant than the quotidian fact that we can build machines that *move faster* than we do, *lift heavier weights* than we do, or *make more accurate measurements* than we do. In other words, digital computing systems and digital technology are artificial, but *not* intelligent in the sense in which we're intelligent; and to that extent, the term "artificial intelligence" is simply an *oxymoron*.

The widely-held yet profoundly false contrary belief—namely, that digital computing systems or digital technology can equal or exceed the essentially embodied innate mental capacities or powers of rational human animals—is what I call *the myth of artificial intelligence*, or "the myth of AI" for short. Tautologically, the myth of AI is a *myth*, with no causal powers of its own. Nevertheless, just as nuclear war technology and biological weapons technology aren't either intelligent or superintelligent, yet still pose an all-too-real dire threat to humankind in the brute, raw sense that their use can kill everyone and destroy the Earth as we know it, so too the myth of AI poses an all-too-real dire threat to humankind's *embodied* or *physical* existence, especially insofar as people *believe that* nuclear weapons technology and biological weapons technology should be controlled by "intelligent" or "superintelligent" digital computers that run or will run drones, killer robots, missiles, and other deadly machines. This "accelerationist" doomsday mindset is especially characteristic of very, very rich technocrats and their scientific lackeys:

> Google cofounder Larry Page thinks superintelligent AI is "just the next step in evolution." In fact, Page, who's worth about $120 billion, has reportedly argued that efforts to prevent AI-driven extinction and protect human consciousness are "speciesist" and "sentimental nonsense."
>
> In July, former Google DeepMind senior scientist Richard Sutton—one of the pioneers of reinforcement learning, a major subfield of AI—said that the technology "could displace us from existence," and that "we should not resist succession." In a 2015 talk, Sutton said, suppose "everything fails" and AI "kill[s] us all"; he asked, "Is it so bad that humans are not the final form of intelligent life in the universe?"
>
> "Biological extinction, that's not the point," Sutton, sixty-six, told me. "The light of humanity and our understanding, our intelligence—our consciousness, if you will—can go on without meat humans." (Lovely, 2024: p. 67)

Moreover, the myth of AI poses a different yet equally dire threat to humankind's cognitive, affective, and practical existence—in short, to humankind's *conscious* and *self-*

conscious existence—even if humankind continues to exist in the embodied or physical sense. Digital computing systems or digital technology that can supposedly equal or exceed rational human animal intelligence are sometimes also called "Artificial General Intelligence" or AGI. Nevertheless, AGI is no more really possible than is Sonny the NS-5 prototype robot in the 2004 movie, *I, Robot*:



A still frame from "I, Robot" (2004, dir. Alex Proyas)

Science fiction and futuristic fantasy are just fine as artistic genres—provided that we don't also start to believe them and act on that belief. But the myth of AI is *pernicious*, since it leads us not only to seriously to *depreciate* and *underestimate* our own mental capacities or powers, but also, by means of our excessive use of and reliance on digital technology, seriously to *neglect* and even *impair* our own mental capacities or powers—this is particularly true in the case of the recent roll-out of Large Language Models (LLMs) or chatbots like ChatGPT: *the invasion of the mind-snatchers* (Hanna, 2023a)—and, to the extent that we knowingly or unknowingly disseminate and perpetuate the myth of AI, also seriously to *misapply* and *misuse* our own mental capacities or powers.

That all being so, in view of the two existential threats posed to humanity by the myth of AI, then the only remaining really important philosophical questions are: (i) why does the myth of AI *persist*?, and (ii*) what it is to be done* about it? I'll offer answers to those hard questions later in this essay. But before we get there, let me now rehearse, briefly but in a step-by-step way, precisely why the myth of AI really *is* a myth, that is, why it is a widely-held yet profoundly *false* belief.

2. Why The Myth of AI Really *is* a Myth

The myth of AI can be explicitly formulated as a two-part philosophical thesis, *the strong thesis of artificial intelligence*, aka strong AI, which says (i) that rational human intelligence can be explanatorily and ontologically reduced to Turing-computable algorithms and the operations of digital computers or digital technology (aka *the thesis of formal mechanism, as it's applied to rational human intelligence*), and (ii) that it's really possible to build digital computing systems or digital technology that are counterpart models of rational human intelligence, such that these systems or technology not only exactly reproduce (aka simulate) all the actual performances of rational human intelligence, but also outperform or surpass it (aka *the counterpart thesis*) (see, e.g., Block, 1980: part 3; Kim, 2011: ch. 6). I'll now describe ten distinct arguments against strong AI.

The first two arguments belong to what I'll call the *old-school*, *phenomenological* critique of strong AI—as originally developed by, for example, Hubert Dreyfus and John Searle in the 1960s, 70s, and 80s.

First, according to Searle's *chinese room argument*, even assuming that a digital computing system or digital technology passes The Turing Test (Turing, 1950), nevertheless it *cannot* have consciousness (or subjective experience) and intentionality (or mental directedness); but *all* human thinking of any kind is conscious and intentional and/or self-conscious, including *all* of its characteristic rational achievements; hence digital computing systems or digital technology *cannot*, even in principle, equal or exceed the characteristic achievements of *conscious intentional* human thinking (Searle, 1980a, 1980b, 1984).

Second, according to Dreyfus's *heideggerian argument*, digital computing systems and digital technology *cannot* engage in unconscious, essentially non-conceptual, non-rule-based, non-inferential, pre-logical, context-sensitive, know-how-driven, skillful, intuitional, sensible activities; but at least *some* human thinking is of this specific kind, including all the characteristic achievements of unconscious affective/emotional, perceptual, and practical, *pre-rational* human thinking; hence AI *cannot*, even in principle, equal or exceed the characteristic achievements of *unconscious pre-rational* human thinking (Dreyfus, 1972, 1979, 1992; Dreyfus and Dreyfus, 1986).

Clearly, however, there's an inconsistency between these two phenomenological lines of argument, since those who use the first line of argument—the "Searlians"—hold that *all* human thinking, including its characteristic rational achievements, is *conscious and intentional*, whereas those who use the second line of argument—the "Dreyfusards" (yes,

it's a pun)—hold that at least *some* of the characteristic achievements of human thinking, including its characteristic pre-rational achievements, are *unconscious*. This inconsistency, in turn, has allowed defenders of strong AI to drive a critical wedge into this unfortunate gap between the Searlians and the Dreyfusards (i) by endlessly *postponing* solving the problem of consciousness, insofar as the strong AI defenders treat it as a "hard" (Chalmers, 1996) and perhaps even insoluble and "mysterian" (McGinn, 1989) problem, but in any case an *open* problem, and (ii) by *also,* at the very same time, using the method of *reverse-engineering*, industriously and industrially designing, building, marketing, and rolling out new and extremely lucrative forms of digital technology, especially those based on "machine learning" systems—neural networks—and robotics, that more and more closely *behaviorally mimic* the characteristic achievements of *un*conscious human thinking.

My *new-school* and *neo-organicist* (Hanna and Paans, 2020; Torday, Miller Jr, and Hanna, 2020) critique of strong AI, however, closes this unfortunate gap in the old, phenomenological Searlian-Dreyfusard critique of AI, by holding, **first**, that consciousness and intentionality are necessarily and completely (i.e., essentially) embodied, yet neither logically nor naturally/nomologically dependent on or reducible to fundamentally material or physical properties, and also **second**, by rejecting the very idea of unconscious human thinking, and by asserting that all of the characteristic pre-rational achievements of human thinking, and indeed all mental activities of any kind, are *also* inherently conscious, even if only *pre-reflectively, non-self-consciously,* and *essentially non-conceptually* conscious (aka "first-order consciousness"): I call this *The Deep Consciousness Thesis* (Hanna and Maiese, 2009; Hanna, 2011a, 2015: esp. section 2.8). If The Essential Embodiment Theory and The Deep Consciousness Thesis are both true, then defenders of strong AI cannot justifiably endlessly postpone solving the problem of consciousness by treating it a hard or mysterian and in any case open problem, and in the meantime, using reverse engineering, create and sell—thereby reaping immense profits—machine-learning-based or robot-based simulations of the characteristic achievements of rational human minded animal thinking, marketing and disseminating them as if they somehow equalled or surpassed the characteristic achievements of human thinking. For essentially embodied deep consciousness constitutes a necessary and indeed essential component of the characteristic *pre-rational* achievements of human thinking, as well as being a necessary and indeed essential component of its characteristic *rational* achievements.

For these reasons, I think that contemporary and future critics of strong AI should rely instead on the following eight new-school and neo-organicist arguments against it.

First, according to *the inadequacy of the turing test argument*, passing The Turing Test, which relies on human judgments about whether a series of written outputs from a digital computing system inside a room manifest intelligence or thinking or not, that's comparable to what's manifested by the written outputs provided by a rational human interlocutor inside another room, clearly is insufficient to show that a digital computing system is intelligent in the sense in which we're intelligent, since human judgers in general are very gullible and therefore easily fooled on all sorts of recognition tests having nothing whatsoever to do with AI—for example, supposedly contacting spirits at séances and other faked mystical experiences—and this gullibility smoothly transfers to The Turing Test; hence the criterion of intelligence presupposed by strong AI is inadequate to capture the presence of intelligence in the sense in which we're intelligent (Hanna, 2023a).

Second, according to *the organicity argument*, rational human mindedness in all its modes, including intelligence and free agency, requires essential embodiment in organic and indeed organismic processes (Hanna and Maiese, 2009; Hanna, 2015, 2018), so since all digital computing systems and digital technology are mechanical and not organic or organismic, then they can't be intelligent in the sense in which we're intelligent (see also Landgrebe and Smith, 2022).

Third, according to *the readability argument*, there are some illegible, meaningless, or nonsensical texts that no digital computing system or digital technology can parse or read, yet rational human animals can indeed parse and read, hence digital computing systems and digital technology cannot be intelligent in the sense in which we're intelligent (Hanna, 2023b).

Fourth, according to *the it's-all-done-with-mirrors argument*, there are some well-specified sets of circumstances in which digital computing systems or digital technology cannot discriminate between left-handed and right-handed but otherwise identical (i.e., enantiomorphic) counterparts, yet rational human animals can indeed discriminate between them, hence digital computing systems and digital technology cannot be intelligent in the sense in which we're intelligent (Hanna, 2023c).

Fifth, according to *the uncomputable functions argument*, digital computing systems or digital technology can't carry out functions or operations in the logico-mathematical sense over domains containing objects or other items that are either non-denumerably finite or non-denumerably infinite, vague, holistic, or entangled, or for which *the rule-following problem* holds, including *the halting problem*, yet rational human animals can indeed perform these very functions or operations, hence digital computing systems and digital technology cannot be intelligent in the sense in which we're intelligent (Hanna, 2023d).

Sixth, according to *the incompleteness argument*, as a specific case of the uncomputable functions argument, no digital computing system or digital technology can carry out functions or operations in the logico-mathematical sense beyond the formal limitations determined by Kurt Gödel's two incompleteness theorems, yet rational human animals can indeed perform these very functions or operations beyond the limits of incompleteness, hence digital computing systems and digital technology cannot be intelligent in the sense in which we're intelligent (Gödel, 1931/1967; Keller, 2023).

Seventh, according to *the babbage's principle argument*, which is a maximally wide-scope generalization of the classical *garbage-in, garbage-out principle*, aka GIGO, no digital computing system or digital technology can perform categorical improvements or upgrades of the intrinsic specific character or quality of the informational inputs, premises, or materials with which they're supplied—for example, from meaninglessness to meaningfulness, or from falsity to truth—yet rational human animals can *creatively transform* these informational inputs, premises, or other materials into categorically improved or upgraded informational outputs, conclusions, or other products, in ten different ways, hence digital computing systems and digital technology cannot be intelligent in the sense in which we're intelligent, especially as regards our authentic human creativity (Hanna, 2023e, 2023f; see also Chomsky, Roberts, and Watamull, 2023).

Eighth, and finally, according to *the "it's a human thing, you wouldn't understand" argument*, our essentially embodied innate mental capacity for *affective intelligence* necessarily transcends the mechanical powers of any and all actual or really possible digital computing systems or digital technology more generally, even when we also fully allow for our finitude, fallibility, and thoroughgoing normative imperfection (Hanna, 2023g).

So, to summarize, there are at least ten old-school or new-school arguments against strong AI: 1. the chinese room argument, 2. the heideggerian argument, 3. the inadequacy of the turing test argument, 4. the organicity argument, 5. the readability argument, 6. the it's-all-done-with-mirrors argument, 7. the uncomputable functions argument, 8. the incompleteness argument, 9. the babbage's principle argument, and 10. the "it's a human thing, you wouldn't understand" argument. Moreover, at least eight of these—i.e., arguments 3. through 10.—are *not* subject to the problems faced by the old-school chinese room argument and heideggerian argument, hence we can confidently conclude that strong AI is false.

3. Why Does The Myth of AI Persist?

I'm now in a position to re-raise the first of two questions I posed at the outset of this essay: given the eight new-school and neo-organicist arguments against strong AI that I just rehearsed—which, when considered as a single collective package, surely has *knockdown* persuasive rational force—and in view of the further fact that the myth of AI has the pernicious consequences I also mentioned at the outset, then *why* does it persist?

I think that the persistence of the myth of AI can be explained by a combination of six essentially *ideological* causal factors: (i) a dogmatic or at least irresponsibly uncritical commitment to the false doctrine of Cartesian dualism, whether substance dualism or property dualism (Hanna and Maiese, 2009), (ii) a dogmatic or at least irresponsibly uncritical commitment to the fantasy of post-humanist or trans-humanist spiritualism (Hanna, 2022, 2023h), (iii) a dogmatic or at least irresponsibly uncritical commitment to the false doctrine of intellectualism about the nature of rational human animals and rational human cognition (Hanna, 2015), (iv) a dogmatic or at least irresponsibly uncritical commitment to false (reductive or non-reductive) materialist/physicalist views about the rational human mind, especially including computational functionalism (Hanna and Maiese, 2009), (v) more generally, a dogmatic or at least irresponsibly uncritical commitment to the false *mechanistic worldview* (Hanna and Paans, 2020), and finally (vi) the hegemony of what I call *the military-industrial-digital complex*[2] (see, e.g., Harper's, 2024), which amasses and reaps immense wealth and political power precisely by means of effectively and relentlessly disseminating and perpetuating the myth of AI, while at the same time hypocritically issuing public warnings about the dangers of runaway AI:

[2] This riffs on a famous phrase in US President Dwight D. Eisenhower's "Farewell Address" in 1961:

> [The] conjunction of an immense military establishment and a large arms industry is new in the American experience. The total influence—economic, political, even spiritual—is felt in every city, every statehouse, every office of the federal government. We recognize the imperative need for this development. Yet we must not fail to comprehend its grave implications. Our toil, resources and livelihood are all involved; so is the very structure of our society. In the councils of government, we must guard against the acquisition of unwarranted influence, whether sought or unsought, by the military–industrial complex. The potential for the disastrous rise of misplaced power exists, and will persist. We must never let the weight of this combination endanger our liberties or democratic processes. We should take nothing for granted. Only an alert and knowledgeable citizenry can compel the proper meshing of the huge industrial and military machinery of defense with our peaceful methods and goals so that security and liberty may prosper together. (See, e.g., Wikipedia, 2024, underlining added)

Yoshua Bengio, fifty-nine, is the second-most cited living scientist, noted for his foundational work on deep learning. Responding to Page and Sutton, Bengio told me, "What they want, I think it's playing dice with humanity's future. I personally think this should be criminalized." A bit surprised, I asked what exactly he wanted outlawed, and he said efforts to build "AI systems that could overpower us and have their own self-interest by design." In May, Bengio began writing and speaking about how advanced AI systems might go rogue and pose an extinction risk to humanity.

Bengio posits that future, genuinely human-level AI systems could improve their own capabilities, functionally creating a new, more intelligent species. Humanity has driven hundreds of other species extinct, largely by accident. He fears that we could be next—and he isn't alone.

Bengio shared the 2018 Turing Award, computing's Nobel Prize, with fellow deep learning pioneers Yann LeCun and Geoffrey Hinton. Hinton, the most cited living scientist, made waves in May when he resigned from his senior role at Google to more freely sound off about the possibility that future AI systems could wipe out humanity. Hinton and Bengio are the two most prominent AI researchers to join the "x-risk" community. Sometimes referred to as AI safety advocates or doomers, this loose-knit group worries that AI poses an existential risk to humanity.

In the same month that Hinton resigned from Google, hundreds of AI researchers and notable figures signed an open letter stating, "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." Hinton and Bengio were the lead signatories, followed by OpenAI CEO Sam Altman and the heads of other top AI labs.

Hinton and Bengio were also the first authors of an October position paper warning about the risk of "an irreversible loss of human control over autonomous AI systems," joined by famous academics like Nobel laureate Daniel Kahneman and Sapiens author Yuval Noah Harari.

LeCun, who runs AI at Meta, agrees that human-level AI is coming but said in a publicdebate against Bengio on AI extinction, "If it's dangerous, we won't build it."

Deep learning powers the most advanced AI systems in the world, from DeepMind's protein-folding model to large language models (LLMs) like OpenAI's ChatGPT. No one really understands how deep learning systems work, but their performance has continued to improve nonetheless. These systems aren't designed to function according to a set of well-understood principles but are instead "trained" to analyze patterns in large datasets, with complex behavior—like language understanding—emerging as a consequence. AI developer Connor Leahy told me, "It's more like we're poking something in a Petri dish"

> than writing a piece of code. The October position paper [see, e.g., Hanna, warns that "no one currently knows how to reliably align AI behavior with complex values."
>
> In spite of all this uncertainty, AI companies see themselves as being in a race to make these systems as powerful as they can—without a workable plan to understand how the things they're creating actually function, all while cutting corners on safety to win more market share. Artificial general intelligence (AGI) is the holy grail that leading AI labs are explicitly working toward. AGI is often defined as a system that is at least as good as humans at almost any intellectual task. It's also the thing that Bengio and Hinton believe could lead to the end of humanity.
>
> Bizarrely, many of the people actively advancing AI capabilities think there's a significant chance that doing so will ultimately cause the apocalypse. A 2022 survey of machine learning researchers found that nearly half of them thought there was at least a 10 percent chance advanced AI could lead to "human extinction or [a] similarly permanent and severe disempowerment" of humanity. Just months before he cofounded OpenAI, Altman said, "AI will probably most likely lead to the end of the world, but in the meantime, there'll be great companies." (Lovely, 2024: pp. 67-68)

For more on this, see section 5 below.

In any case, the first five factors are classically philosophical factors, so they're at least in principle open to critical rational argumentation, refutation, and correction. But the same isn't the case with respect to the sixth factor, which is essentially social-institutional and political in nature. Sadly, however, I think it's more than merely reasonable to hold that the hegemony of the military-industrial-digital complex is the *principal* cause of the persistence of the myth of AI, as Garrison Lovely cogently and crisply puts it:

> The debate playing out in the public square may lead you to believe that we have to choose between addressing AI's immediate harms and its inherently speculative existential risks. And there are certainly trade-offs that require careful consideration.
>
> But when you look at the material forces at play, a different picture emerges: in one corner are trillion-dollar companies trying to make AI models more powerful and profitable; in another, you find civil society groups trying to make AI reflect values that routinely clash with profit maximization.
>
> In short, it's capitalism versus humanity. (Lovely, 2024: p. 79)

Nevertheless, radical social-institutional and political change for the better, or even the best, is in fact really possible, by means of what Michelle Maiese and I have called *the*

mind-body politic and *the enactive-transformative principle* (Maiese and Hanna, 2019). So to end this section on an upbeat note, there's at least *some* rational hope for debunking the myth of AI, by devolving and enactively transforming the military-industrial-digital complex.

4. Oppenheimer, Kaczynski, Shelley, and Frankenscience

> What we are creating now [with the atomic bomb] is a monster whose influence is going to change history, provided there is any history left, yet it would be impossible not to see it through, not only for the military reasons, but it would also be unethical from the point of view of the scientists not to do what they know is feasible, no matter what terrible consequences it may have. (John von Neumann, as reported by his wife and quoted in [Dyson, 2012: p. 62].)

What do J. Robert Oppenheimer (1904-1967) and Theodore Kaczynski (1942-2023) have in common?[3] They were both exceptionally intelligent, indeed brilliant, and sometimes troubled, people. They both held BAs from Harvard and received their PhDs very early, at the ages of 23 and 25 respectively. They were both formal scientists: a theoretical physicist and a mathematician respectively. They both taught at UC-Berkeley. Oppenheimer was the father of the atomic bomb, aka the A-bomb, which killed at least 110,000 people at Hiroshima and Nagasaki in 1945, bringing about Japan's unconditional surrender and ending World War II. And Kaczynski was the father of the unabomb (a neologism deriving from an FBI case file labelled "university and airline bomber"), a letter bomb campaign against technocrats that killed three people and injured 23 others between 1978 and 1995.

There are also important dissimilarities. Oppenheimer was never accused of or charged with war crimes, and indeed, until his run-in with McCarthyism in the 1950s, he was universally regarded as a national public hero of science. Kaczynski, by contrast, was arrested in 1996, sentenced to life imprisonment without parole in 1998, and ultimately committed suicide in prison in June 2023. But most importantly of all, at least during World War II, Oppenheimer *fervently affirmed*, whereas from the 1970s onwards Kaczynski *violently opposed*, what I'll call *the thesis of frankenscience*, formulated by another Manhattan Project member, John von Neumann, as per the epigraph at the top of this section. For the purposes of this essay, I'll slightly reformulate the thesis of frankenscience as follows: *scientists are always obligated, as scientists, to do whatever they know is feasible, no*

[3] For more information about Oppenheimer and Kaczynski, see these two excellent documentaries: (Else, 1980; Dammbeck, 2003).

matter how morally objectionable this is and no matter what bad consequences this may have for humankind.

The neologism "frankenscience" is obviously a riff on the title of Mary Shelley's brilliant and worldwide-famous 1818 novella, *Frankenstein; Or, the Modern Prometheus*, which of course is about a scientist, Victor Frankenstein, who creates a humanoid from corpses and brings it to life with electricity, only to discover that he's created a monster. In turn, Frankenstein's recognition of his tragic mistake can be taken to be Shelley's direct rejection of frankenscience:

> Learn from me, if not by my precepts, at least by my example, how dangerous is the acquirement of [scientific] knowledge, and how much happier that man is who believes his native town to be the world, than he who aspires to become greater than his nature will allow. (Shelley, 1818: vol. 1, ch. 3)

I strongly agree with Shelley. It's crucial to recognize, however, that it's entirely possible to be *pro-science*, but also *anti-frankenscience*, and indeed that's my view. Much science is intrinsically valuable, morally unobjectionable, and also has good consequences for humankind. But scientists are *never* obligated, as scientists, to do whatever they know is feasible, *whenever* this is morally objectionable or has bad consequences for humankind. On the contrary, they're always morally obligated to *refuse* to participate in scientific research of these kinds and also to *oppose* the use of science for such purposes, for the sake of sufficient respect for human dignity. Correspondingly, as I said, I strongly agree with Shelley's direct rejection of frankenscience, whatever her views might have been about science per se. In any case, in sections 5 to 7 below I want to extend that direct rejection of frankenscience to the contemporary research program of so-called artificial intelligence or AI.

5. Hinton & Me

In an open letter published online in late March 2023, directed not only to the digital technology and AI community in particular but also to the world more generally, card-carrying members of the military-industrial-digital complex—including Elon Musk and Steve Wozniak—urged "all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4" (Future of Life Institute, 2023). Here's their letter in full, leaving out the Notes and References:

> AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research and acknowledged by top AI labs. As stated in the widely-endorsed Asilomar AI Principles, *Advanced AI could represent a profound change*

in the history of life on Earth, and should be planned for and managed with commensurate care and resources. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one—not even their creators—can understand, predict, or reliably control.

Contemporary AI systems are now becoming human-competitive at general tasks, and we must ask ourselves: *Should* we let machines flood our information channels with propaganda and untruth? *Should* we automate away all the jobs, including the fulfilling ones? *Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? *Should* we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders. **Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable.** This confidence must be well justified and increase with the magnitude of a system's potential effects. OpenAI's recent statement regarding artificial general intelligence, states that *"At some point, it may be important to get independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of compute used for creating new models."* We agree. That point is now.

Therefore, **we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4**. This pause should be public and verifiable, and include all key actors. If such a pause cannot be enacted quickly, governments should step in and institute a moratorium.

AI labs and independent experts should use this pause to jointly develop and implement a set of shared safety protocols for advanced AI design and development that are rigorously audited and overseen by independent outside experts. These protocols should ensure that systems adhering to them are safe beyond a reasonable doubt. This does *not* mean a pause on AI development in general, merely a stepping back from the dangerous race to ever-larger unpredictable black-box models with emergent capabilities.

AI research and development should be refocused on making today's powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal.

In parallel, AI developers must work with policymakers to dramatically accelerate development of robust AI governance systems. These should at a minimum include: new and capable regulatory authorities dedicated to AI; oversight and tracking of highly capable AI systems and large pools of computational capability; provenance and watermarking systems to help distinguish real from synthetic and to track model leaks; a robust auditing and certification ecosystem; liability for AI-caused harm; robust public funding for technical AI safety research; and well-resourced institutions for coping with

> the dramatic economic and political disruptions (especially to democracy) that AI will cause.
>
> Humanity can enjoy a flourishing future with AI. Having succeeded in creating powerful AI systems, we can now enjoy an "AI summer" in which we reap the rewards, engineer these systems for the clear benefit of all, and give society a chance to adapt. Society has hit pause on other technologies with potentially catastrophic effects on society. We can do so here. Let's enjoy a long AI summer, not rush unprepared into a fall. (Future of Life Institute, 2023)

I *strongly disagree* with what's being asserted and argued in this letter. For people with a huge vested interest in designing, producing, and marketing digital technology and AI, to call on AI labs and the fabulously wealthy global technocratic capitalist corporations (including the corporations they own or founded, for example, Musk/Tesla and Wozniak/Apple) that design, produce, and market digital technology, to "pause" their research on Large Language Models (LLMs) and the chatbots that implement these models "beyond GPT-4" for "at least six months,"[4] *and then* go forward again full-steam-ahead, having paid public lip-service to "digital ethics," so that these corporations can create an all-out existential attack on humankind, is just like it would have been for leading US generals during the latter phases of World War II to call on The Manhattan Project to "pause" their work on the atomic bomb for "at least six months," then go forward again full-steam-ahead, having paid public lip-service to "just war theory," so that the USA could drop the A-bomb not just once but twice on hundreds of thousands of Japanese civilians, and thereby create a permanent existential threat for humankind.

On 1 May 2023—significantly, *May Day*—Geoffrey Hinton, a groundbreaking researcher on neural networks and LLMs, quit his job at Google and then publicly made a very similar point, also explicitly using the analogy with The Manhattan Project:

> As companies improve their A.I. systems, [Hinton] believes, they become increasingly dangerous. "Look at how it was five years ago and how it is now," he said of A.I. technology. "Take the difference and propagate it forwards. That's scary."

[4] The authors and signatories of the open letter say that "[i]f such a [six-months-minimum] pause cannot be enacted quickly, governments should step in and institute a moratorium" (Future of Life Institute, 2023). Now, "moratorium" means "temporary prohibition of an activity." So, supposing that a self-adopted pause or government moratorium were to occur, then this still logically or at least conversationally implies that sooner or later, giant AI experiments and LLM/chatbot technology will be resumed. On the contrary, I'm saying that it should be shut down permanently in its infancy, just as A-bomb research and nuclear weapons technology should have been shut down permanently in its infancy.

Until last year, he said, Google acted as a "proper steward" for the technology, careful not to release something that might cause harm. But now that Microsoft has augmented its Bing search engine with a chatbot — challenging Google's core business — Google is racing to deploy the same kind of technology. The tech giants are locked in a competition that might be impossible to stop, Dr. Hinton said.

His immediate concern is that the internet will be flooded with false photos, videos and text, and the average person will "not be able to know what is true anymore."
He is also worried that A.I. technologies will in time upend the job market. Today, chatbots like ChatGPT tend to complement human workers, but they could replace paralegals, personal assistants, translators and others who handle rote tasks. "It takes away the drudge work," he said. "It might take away more than that."

Down the road, he is worried that future versions of the technology pose a threat to humanity because they often learn unexpected behavior from the vast amounts of data they analyze. This becomes an issue, he said, as individuals and companies allow A.I. systems not only to generate their own computer code but actually run that code on their own. And he fears a day when truly autonomous weapons—those killer robots—become reality.

"The idea that this stuff could actually get smarter than people—a few people believed that," he said. "But most people thought it was way off. And I thought it was way off. I thought it was 30 to 50 years or even longer away. Obviously, I no longer think that."

Many other experts, including many of his students and colleagues, say this threat is hypothetical. But Dr. Hinton believes that the race between Google and Microsoft and others will escalate into a global race that will not stop without some sort of global regulation.

But that may be impossible, he said. Unlike with nuclear weapons, he said, there is no way of knowing whether companies or countries are working on the technology in secret. The best hope is for the world's leading scientists to collaborate on ways of controlling the technology. "I don't think they should scale this up more until they have understood whether they can control it," he said.

Dr. Hinton said that when people used to ask him how he could work on technology that was potentially dangerous, he would paraphrase Robert Oppenheimer, who led the U.S. effort to build the atomic bomb: "When you see something that is technically sweet, you go ahead and do it."

He does not say that anymore. (New York Times, 2023)

What Hinton (apparently) and I (absolutely) strongly believe, then, is that *we ought to ban all giant AI experiments and LLM/chatbot technology while they are still in their infancy*, just as *we ought to have banned all atomic bomb experiments and nuclear weapons technology while they were still in their infancy.* My strong belief, in turn, is a direct expression and logical implication of the doctrine I call *dignitarian neo-luddism with respect to digital technology*, which I'll now explain and justify.

6. Dignitarian Neo-Luddism With Respect to Digital Technology

For clarity's sake, I'll start with some definitions. According to *The Oxford Encyclopedic Dictionary*, the term "Luddite" means this:

> 1. *hist.* a member of the bands of English craftsmen who, when their jobs were threatened by the progressive introduction of machinery into their trades in the early 19th c. attempted to reverse the trend towards mechanization by wrecking the offending machines…. 2. a person opposed to increased industrialization or new technology …. [perh. f. Ned *Lud[d],* an insane person said to have destroyed two stocking-frames *c.* 1799] (Hawkins and Allen, 1991: p. 856)

Generalizing from that, and also precisifying a little, I'll say that classical or modern *Luddism* says that

> all mechanical technology is bad and wrong, because it harms and oppresses ordinary people (i.e., people other than technocrats), and therefore all mechanical technology should be rejected and destroyed, and this may also include using violence or other terrorist tactics against technocrats.

For example, Ted Kaczynski was a modern Luddite and indeed also a terrorist modern Luddite.

By an important contrast, however, as I'm understanding it, modern *neo-Luddism*—see, for example, Glendinning, 1990—says that

> not all mechanical technology is bad and wrong, but instead all and only the mechanical technology that harms and oppresses ordinary people (i.e., people other than technocrats) is bad and wrong, and therefore all and only this bad and wrong mechanical technology should be rejected but not—except in extreme cases of mechanical technology whose coercive use is actually violently harming and oppressing ordinary people, for example, weapons being used for mass

> destruction or mass murder—destroyed, rather only either simply refused, non-violently dismantled, or radically transformed into its moral opposite.

Now, by *digital technology* I mean all mechanical technology that inherently involves computers, algorithms, digital data or information, so-called artificial intelligence/AI, or robotics. Then, *neo-Luddism with respect to digital technology* says that

> not all digital technology is bad and wrong, but instead all and only the digital technology that harms and oppresses ordinary people (i.e., people other than digital technocrats) is bad and wrong, and therefore all and only this bad and wrong digital technology should be rejected but not—except in extreme cases of digital technology whose coercive use is actually violently harming and oppressing ordinary people, for example, digitally-driven weapons or weapons-systems being used for mass destruction or mass murder—destroyed, rather only either simply refused, non-violently dismantled, or radically transformed into its moral opposite.

Finally, *dignitarian neo-Luddism with respect to digital technology* says that

> not all digital technology is bad and wrong, but instead all and only the digital technology that harms and oppresses ordinary people (i.e., people other than digital technocrats), by either failing to respect our human dignity sufficiently or by outright violating our human dignity, is bad and wrong, and therefore all and only this bad and wrong digital technology should be rejected but not—except in extreme cases of digital technology whose coercive use is actually violently harming and oppressing ordinary people, for example, digitally-driven weapons or weapons-systems being used for mass destruction or mass murder—destroyed, rather only either simply refused, non-violently dismantled, or radically transformed into its moral opposite.

It needs to emphasized and re-emphasized that just as I'm *pro-science* while also being *anti-frankenscience*, so too dignitarian neo-Luddism with respect to digital technology is *also* committed to the *positive* dignitarian moral doctrine that *some* digital technology is permissible and good, and therefore *ought to be used*, precisely because it's morally unobjectionable and promotes the betterment of humankind, and more generally sufficiently respects human dignity. For example, in my opinion this is true of posting or self-publishing essays about dignitarian digital/AI ethics for universal free sharing on the internet. Why else would I be doing it? But in the context of this essay, I'm focusing primarily on the negative dignitarian moral doctrine.

So much for the definitions, and now for some moral imperatives. What I strongly believe is that *we all ought to be dignitarian neo-Luddites with respect to digital technology*. Why? To be sure, there are many ways in which digital technology can be bad and wrong in the dignitarian sense, including invasive digital surveillance, digitally-driven weapons and weapon systems, algorithmic bias, and digital manipulation and nudging. And of course there are also ways in which digital technology can be bad and wrong in the utilitarian sense, for example, putting many people out of work. But the principal reason for being a dignitarian neo-Luddite with respect to digital technology is that *our excessive use of and indeed addiction to digital technology is systematically undermining our innate capacities for thinking, caring, and acting for ourselves*. This is *preeminently* true with respect to the new chatbots—for example, ChatGPT and LaMDA—and what, in sections 1 to 3 above, I've called *the myth of AI* more generally (see, e.g., Hanna, 2023a, 2023b, 2023c, 2023d, 2023e, 2023f), but also *to an increasingly important degree* true for our excessive use of and addiction to smart-phones, desktop and laptop computers, the internet, social media, and so-on and so-forth. When you combine our excessive use of and addiction to chatbots and AI with our excessive use of and addiction to smart-phones, desktop and laptop computers, the internet, social media, etc., the result is nothing less than *an all-out existential attack on our rational human mindedness or intelligence.*

By "our rational human mindedness or intelligence" I mean the essentially embodied, unified set of basic innate cognitive, affective, and practical capacities present in all and only those human animals possessing the essentially embodied neurobiological basis of those capacities, namely: (i) *consciousness*, i.e., subjective experience, (ii) *self-consciousness*, i.e., consciousness of one's own consciousness, second-order consciousness, (iii) *caring*, i.e., desiring, emoting, or feeling, (iv) *sensible cognition*, i.e., sense-perception, memory, or imagination, (v) *intellectual cognition*, i.e., conceptualizing, believing, judging, or inferring, (vi) *volition*, i.e., deciding, choosing, or willing, and (vii) *free agency*, i.e., free will and practical agency. This unified set of capacities constitutes our *human real personhood*, which in turn is *the metaphysical ground of our human dignity* (Hanna, 2023i, 2023j). Therefore, this all-out existential attack on our rational human mindedness or intelligence is also an all-out existential attack on our human dignity.

The Cassandra-like prophecy and warning that I'm issuing, however, is *not* that chatbots or AI more generally could ever become rational, super-intelligent, and morally satanic, and then run amok. Interestingly, that seems to be the main source of Bengio's and Hinton's worries. But in fact, it's *metaphysically impossible* for computing machines ever to be rationally minded or intelligent in the sense that *we're* rationally minded or intelligent, because (i) it's *metaphysically necessary* that all creatures possessing the seven basic innate capacities I listed above are complex living organisms, i.e., *animals* (Hanna and Maiese, 2009), hence not *machines*, hence not *computing machines*, and (ii) it's also

metaphysically necessary that our rational mindedness or intelligence includes (iia) an innate non-basic ("non-basic," in the sense that it essentially depends on the seven basic innate capacities listed in the just-previous paragraph) capacity for *spontaneous creativity*, and also (iib) an innate non-basic capacity for either conceptual or essentially non-conceptual *a priori intuition* of (iib1) innately-specified universal, unconditional, a priori or non-empirical moral principles such as *everyone ought always to choose and act with sufficient respect for everyone's dignity, including their own*, (iib2) universal, unconditional, a priori or non-empirical logical principles such as the minimal principle of non-contradiction, namely, *not every statement is both true and false*, and (iib3) the universal a priori or non-empirical formal structures of the orientable, three-dimensional space and the forward-directed, processual, purposive, asymmetric organic time in which our minded animal bodies are ineluctably embedded (Hanna, 2006, 2015, 2018), *none of which can ever exist in computing machinery*. And, closely related to these metaphysically necessary modal facts, there are also various other linguistic, logical, mathematical, and metaphysical reasons why computing machinery can never be rationally minded or intelligent in the sense that *we're* rationally minded or intelligent, as per the ten arguments I presented in section 2 above (see also Chomsky, Roberts, and Watamull, 2023; Hanna, 2023a, 2023b, 2023c, 2023d, 2023e, 2023f, 2023g; Keller, 2023; Landgrebe and Smith, 2022).

On the contrary, the Cassandra-like prophecy and warning that I'm issuing about this all-out existential attack on our rational human mindedness or intelligence is instead directed at *the global technocratic capitalist corporations—especially those that supply weapons and surveillance systems for military and government use—millionaires, and billionaires* who reap immense profits and wield immense political power by designing, producing, marketing, and above all *controlling* our use of and reliance on digital technology: namely, the members of the military-industrial-digital complex. Correspondingly, my Cassandra-like prophecy and warning is simply this:

> the members of the military-industrial-digital complex are systematically harming and oppressing ordinary people like us by not only enabling but also effectively mandating our excessive use of and addiction to digital technology, which in turn systematically undermines our innate capacities for thinking, caring, and acting for ourselves, and therefore undermines our human real personhood, and thereby violates our human dignity—therefore, we ought to ban all giant AI experiments and LLM/chatbot technology while they are still in their infancy, just as we ought to have banned all atomic bomb experiments and nuclear weapons technology while they were still in their infancy.

7. What is to be Done?

Now, to address the second question I raised at the outset of this essay, what is to be done? Perhaps dignitarian neo-Luddism with respect to digital technology will become a worldwide, world-changing social and political movement, comparable to the Ban-the-Bomb and anti-nuclear movement: I wholeheartedly hope so. Indeed, as of February 2024, there are some positive indications (Guardian, 2024; Lovely, 2024). If we ban all further giant AI experiments and LLM/chatbot technology right now, when it's already obvious what their existential threat to humankind is, then the world will be a substantially better place, just as the world would have been a substantially better place if we had banned the A-bomb and nuclear weapons technology immediately after its initial test on 16 July 1945, when it was already obvious what their existential threat to humankind would be. Indeed, in a film-interview late in his life, Oppenheimer was asked what he thought about Senator Robert F. Kennedy's efforts to urge President Lyndon Johnson to start talks about preventing the development and spread of nuclear weapons. Oppenheimer replied: "It's 20 years too late… It should have been done the day after Trinity" (Else, 1980). It's now "the day after ChatGPT," and therefore we should be acting accordingly.

But, in the meantime, given the immense power of the military-industrial-digital complex, individual dignitarian digital neo-Luddites like you and I (assuming that you agree with what I've argued so far, that is) cannot do very much to change the world. Nevertheless, and now updating J.C. Scott's "weapons of the weak" (Scott, 1985) for our 21st century context, a world pervaded by digital technology and dominated by the military-industrial-digital complex, I do think that we individually *can* become what I'll call *daily dignitarian digital refusards,* aka DDDRs.[5] How? We can become DDDRs by *resolving to cultivate our own innate capacities in a self-disciplined and autonomous way, for six waking hours every day, altogether independently of digital technology, to the extent that this is humanly possible*.

To take just a few of many examples of daily dignitarian digital refusardism, you can *log off all your digital devices for six waking hours a day*, you can *refuse to use ChatGPT or any other chatbot altogether*, and you can instead do some or all of these things: read hard-copy books; write out your ideas and thoughts longhand (Van der Weel and Van der Meer, 2024); memorize and recite poetry; memorize things about subject areas that especially interest you; do mental calculations; do logic puzzles, acrostics, crossword puzzles, etc.; go for long contemplative walks; sit in a park for an hour or so; work in a garden; look carefully at the natural landscape around you; meditate for fifteen minutes

[5] Pronounced "didd-ers."

or half an hour; cook, eat, and drink; doodle or draw longhand, or paint; play music on a real musical instrument, whistle, or sing; reminisce; get together with your loved ones or friends in person and talk about anything under the sun *other than* what's being enabled or mandated by the members of the military-industrial-digital complex; and fall or re-fall in love with someone or something.

Even the military-industrial-digital complex, with all its immense power, *can't* stop you from practicing daily dignitarian digital refusardism in its many modes, each of which involves the self-disciplined, autonomous cultivation of your innate capacities. And if the day does ever come when the military-industrial-digital complex *can* stop you from being a DDDR, whether by means of permanently-implanted brain-computer interfaces or some other malign instruments of digital-technological harm and oppression, then that will really-&-truly be the end of the road for rational humankind and human dignity, even if humankind *weren't* wiped out by the first existential threat of the myth of AI, the threat to our embodied or physical existence. For that will be the day that frankenscience took control of our conscious and self-conscious existence.[6, 7]

[6] For a similar line of thinking, see also (Corbyn, 2023; Farhany, 2023).

[7] I'm grateful to Elma Berisha, Scott Heftler, Michelle Maiese, and Otto Paans for thought-provoking conversations or correspondence on and around the main topics of this essay; to Donald Stanley for drawing my attention to (Lovely, 2024); and to Jack Leissring for drawing my attention to (Harper's, 2024).

REFERENCES

(Block, 1980). Block, N. (ed.), *Readings in the Philosophy of Psychology*. 2 vols., Cambridge, MA: Harvard Univ. Press. Vol. 1.

(Chalmers, 1996). Chalmers, D., *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford Univ. Press.

(Chomsky, Roberts, and Watumull, 2023). Chomsky, N., Roberts, I. and Watumull, J. "Noam Chomsky: The False Promise of ChatGPT." *New York Times*. 8 March. Available online at URL = <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.

(Corbyn, 2023). Corbyn, Z. "Prof Nita Farahany: 'We Need a New Human Right to Cognitive Liberty'." *The Guardian*. 4 March. Available online at URL = <https://www.theguardian.com/science/2023/mar/04/prof-nita-farahany-we-need-a-new-human-right-to-cognitive-liberty>.

(Dammbeck, 2003). Dammbeck, L. (dir) *The Net: The Unabomber, LSD, and the Internet* (*Das Netz*). Available online at URL = <https://archive.org/details/DasNetz_LutzDammbeck>.

(Dreyfus, 1972). Dreyfus, H. *What Computers Can't Do*. Cambridge MA: MIT Press.

(Dreyfus, 1979). Dreyfus, D. *What Computers Can't Do*. 2nd edn., Cambridge MA: MIT Press.

(Dreyfus, 1992). Dreyfus, H. *What Computers Still Can't Do.* Cambridge MA: MIT Press.

(Dreyfus and Dreyfus, 1986). Dreyfus, H. and Dreyfus, S. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Oxford: Blackwell.

(Dyson, 2012). Dyson, D. *Turing's Cathedral*. New York: Vintage.

(Else, 1980). Else, J. (dir.) *The Day After Trinity: J. Robert Oppenheimer and the Atomic Bomb*.

(Farhany, 2023). Farhany, N. *The Battle for the Brain*. New York: St Martin's Press.

(Future of Life Institute, 2023). "Pause Giant AI Experiments: An Open Letter." *Future of Life Institute*. 22 March. Available online at URL = <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

(Guardian, 2024). Lamont, T. " 'Humanity's remaining timeline? It looks more like five years than 50': Meet the Neo-Luddites Warning of an AI Apocalypse." *The Guardian*. 17 February. Available online at URL = <https://www.theguardian.com/technology/2024/feb/17/humanitys-remaining-timeline-it-looks-more-like-five-years-than-50-meet-the-neo-luddites-warning-of-an-ai-apocalypse>.

(Glendinning, 1990). Glendinning, C. "Notes toward a Neo-Luddite Manifesto." *The Anarchist Library*. Available online at URL = <https://theanarchistlibrary.org/library/chellis-glendinning-notes-toward-a-neo-luddite-manifesto>.

(Gödel, 1931/1967). Gödel, K. "On Formally Undecidable Propositions of Principia Mathematica and Related Systems." In J. van Heijenoort (ed.), *From Frege to Gödel.* Cambridge MA: Harvard Univ. Press. Pp. 596-617.

(Hanna, 2006). Hanna, R. *Rationality and Logic*. Cambridge MA: MIT Press. Also available online in preview at URL = <https://www.academia.edu/21202624/Rationality_and_Logic>.

(Hanna, 2011). Hanna, R. "Minding the Body." *Philosophical Topics* 39: 15-40. Available online in preview at URL = <https://www.academia.edu/4458670/Minding_the_Body>.

(Hanna, 2015). Hanna, R. *Cognition, Content, and the A Priori: A Study in the Philosophy of Mind and Knowledge.* THE RATIONAL HUMAN CONDITION, Vol. 5. Oxford: Oxford Univ. Press. Also available online in preview at URL = <https://www.academia.edu/35801833/The_Rational_Human_Condition_5_Cognition_Content_and_the_A_Priori_A_Study_in_the_Philosophy_of_Mind_and_Knowledge_OUP_2015_>.

(Hanna, 2018). Hanna, R. *Deep Freedom and Real Persons: A Study in Metaphysics.* THE RATIONAL HUMAN CONDITION, Vol. 2. New York: Nova Science. Available online in preview at URL = <https://www.academia.edu/35801857/The_Rational_Human_Condition_2_Deep_Freedom_and_Real_Persons_A_Study_in_Metaphysics_Nova_Science_2018_>.

(Hanna, 2022). Hanna, R. "Turing, Strong AI, and The Fantasy of Transhumanist Spiritualism." Unpublished MS. Available online at URL = <https://www.academia.edu/82130775/Turing_Strong_AI_and_The_Fantasy_of_Transhumanist_Spiritualism_June_2022_version_>.

(Hanna, 2023a). Hanna, R. "How and Why ChatGPT Failed The Turing Test." Unpublished MS. Available online at URL = <https://www.academia.edu/94870578/How_and_Why_ChatGPT_Failed_The_Turing_Test_January_2023_version_>.

(Hanna, 2023b). Hanna, R. "Are There Some Legible Texts That Even The World's Most Sophisticated Robot Can't Read?" Unpublished MS. Available online at URL = <https://www.academia.edu/95866304/Are_There_Some_Legible_Texts_That_Even_The_Worlds_Most_Sophisticated_Robot_Cant_Read_January_2023_version_>.

(Hanna, 2023c). Hanna, R. "It's All Done With Mirrors: A New Argument That Strong AI is Impossible." Unpublished MS. Available online at URL = <https://www.academia.edu/95296914/Its_All_Done_With_Mirrors_A_New_Argument_That_Strong_AI_is_Impossible_January_2023_version_>.

(Hanna, 2023d). Hanna, R. "How and Why to Perform Uncomputable Functions." Unpublished MS. Available online at URL = <https://www.academia.edu/87165326/How_and_Why_to_Perform_Uncomputable_Functions_March_2023_version_>.

(Hanna, 2023e). Hanna, R. "Babbage-In, Babbage-Out: On Babbage's Principle." Unpublished MS. Available online at URL = <https://www.academia.edu/101462742/Babbage_In_Babbage_Out_On_Babbages_Principle_May_2023_version_>.

(Hanna, 2023f). "Necessary and Sufficient Conditions for Authentic Human Creativity." Unpublished MS. Available online at URL = <https://www.academia.edu/101633470/Necessary_and_Sufficient_Conditions_for_Authentic_Human_Creativity_May_2023_version_>.

(Hanna, 2023g). Hanna, R. "'It's a Human Thing. You Wouldn't Understand.' Computing Machinery and Affective Intelligence." Unpublished MS. Available online at URL = <https://www.academia.edu/102664604/_Its_a_Human_Thing_You_Wouldnt_Understand_Computing_Machinery_and_Affective_Intelligence_June_2023_version_>.

(Hanna, 2023h). Hanna, R. "Essentially Embodied Kantian Selves and The Fantasy of Transhuman Selves," *Studies in Transcendental Philosophy* 3. Available online at URL = <https://ras.jes.su/transcendental/s271326680021060-6-1-en>, and also in preview at URL = <https://www.academia.edu/93343682/Essentially_Embodied_Kantian_Selves_and_The_Fantasy_of_Transhuman_Selves_Forthcoming_in_Studies_in_Transcendental_Philosophy_3_2023_>.

(Hanna, 2023i). Hanna, R. "Dignity, Not Identity." Unpublished MS. Available online at URL = <https://www.academia.edu/96684801/Dignity_Not_Identity_February_2023_version_>.

(Hanna, 2023j). Hanna, R. "In Defence of Dignity." *Borderless Philosophy* 6: 77-98. Available online at URL = <https://www.cckp.space/single-post/bp6-2023-robert-hanna-in-defence-of-dignity-77-98>.

(Hanna and Maiese, 2009). Hanna, R. and Maiese, M. *Embodied Minds in Action*. Oxford: Oxford Univ. Press. Available online in preview at URL = <https://www.academia.edu/21620839/Embodied_Minds_in_Action>.

(Hanna and Paans, 2020). Hanna, R. and Paans, O. "This is the Way the World Ends: A Philosophy of Civilization Since 1900, and A Philosophy of the Future." *Cosmos & History* 16, 2: 1-53. Available online at URL = <https://cosmosandhistory.org/index.php/journal/article/view/865>.

(Harper's, 2024). Cockburn, A. "The Pentagon's Silicon Valley Problem." *Harper's Magazine*. March. Available online at URL = <https://harpers.org/archive/2024/03/the-pentagons-silicon-valley-problem-andrew-cockburn/>.

(Hawkins and Allen, 1991). Hawkins, J.M. and Allen, R. (eds.), *The Oxford Encyclopedic English Dictionary*. Oxford: Clarendon/Oxford Univ. Press.

(Keller, 2023). Keller, A. "Artificial, But Not Intelligent: A Critical Analysis of AI and AGI." *Against Professional Philosophy*. 5 March. Available online at URL = <https://againstprofphil.org/2023/03/05/artificial-but-not-intelligent-a-critical-analysis-of-ai-and-agi/>.

(Kim, 2011). Kim, J. *Philosophy of Mind*. 3rd edn., Boulder, CO: Westview.

(Landgrebe and Smith, 2022). Landgrebe, J. and Smith, B. *Why Machines Will Never Rule the World: Artificial Intelligence without Fear*. London: Routledge.

(Lovely, 2024). Lovely, G. "Can Humanity Survive AI?" *Jacobin* 52. Winter. Pp. 66-79. Also available online at URL = <https://jacobin.com/2024/01/can-humanity-survive-ai/>.

(Maiese and Hanna, 2019). Maiese, M. and Hanna, R. *The Mind-Body Politic*. London: Palgrave Macmillan. Available online in preview at URL = <https://www.academia.edu/38764188/The_Mind-Body_Politic_Preview_Co-authored_with_Michelle_Maiese_forthcoming_from_Palgrave_Macmillan_in_July_2019_>.

(McGinn, 1989). McGinn, C. "Can We Solve the Mind-Body Problem?" *Mind* 98: 349-366.

(New York Times, 2023). Metz, C. "'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead." *The New York Times*. 1 May. Available online at URL = <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>.

(RAND, 2023). Price, C.C. and Woods, M. "Tackling the Existential Threats from Artificial Intelligence." 11 July. *TheRANDBlog*. Available online at URL = <https://www.rand.org/pubs/commentary/2023/07/tackling-the-existential-threats-from-artificial-intelligence.html>.

(Scott, 1985). Scott, J.C. *Weapons of the Weak: Everyday Forms of Peasant Resistance*. New Haven CT: Yale Univ. Press.

(Searle, 1980a). Searle, J. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3: 417-424.

(Searle, 1980b). Searle, J. "Intrinsic Intentionality." *Behavioral and Brain Sciences* 3: 450-456

(Searle, 1984). Searle, J. *Minds, Brains, and Science*. Cambridge, MA: Harvard Univ. Press.

(Shelley, 1818). Shelley, M. *Frankenstein; Or, the Modern Prometheus*. Available online at URL = <http://www.rc.umd.edu/editions/frankenstein>.

(Torday, Miller Jr, and Hanna, 2020). Torday, J., Miller Jr, W.B., and Hanna, R., "Singularity, Life, and Mind: New Wave Organicism." In J. Torday and W.B. Miller Jr, *The Singularity of Nature*. Cambridge: Royal Society of Chemistry. Ch. 20. Pp. 206-246.

(Turing, 1950). Turing, A. "Computing Machinery and Intelligence." *Mind* 59: 433–460.

(Van der Weel and Van der Meer, 2024). Van der Weel, F.R. and Van der Meer, A.L.H. "Handwriting But Not Typewriting Leads to Widespread Brain Connectivity: A High-Density EEG Study with Implications for the Classroom." *Frontiers in Psychology* 14. 25 January. Available online at URL = <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1219945/full>.

(Wikipedia, 2024). Wikipedia. "Military-Industrial Complex." Available online at URL = <https://en.wikipedia.org/wiki/Military%E2%80%93industrial_complex>.